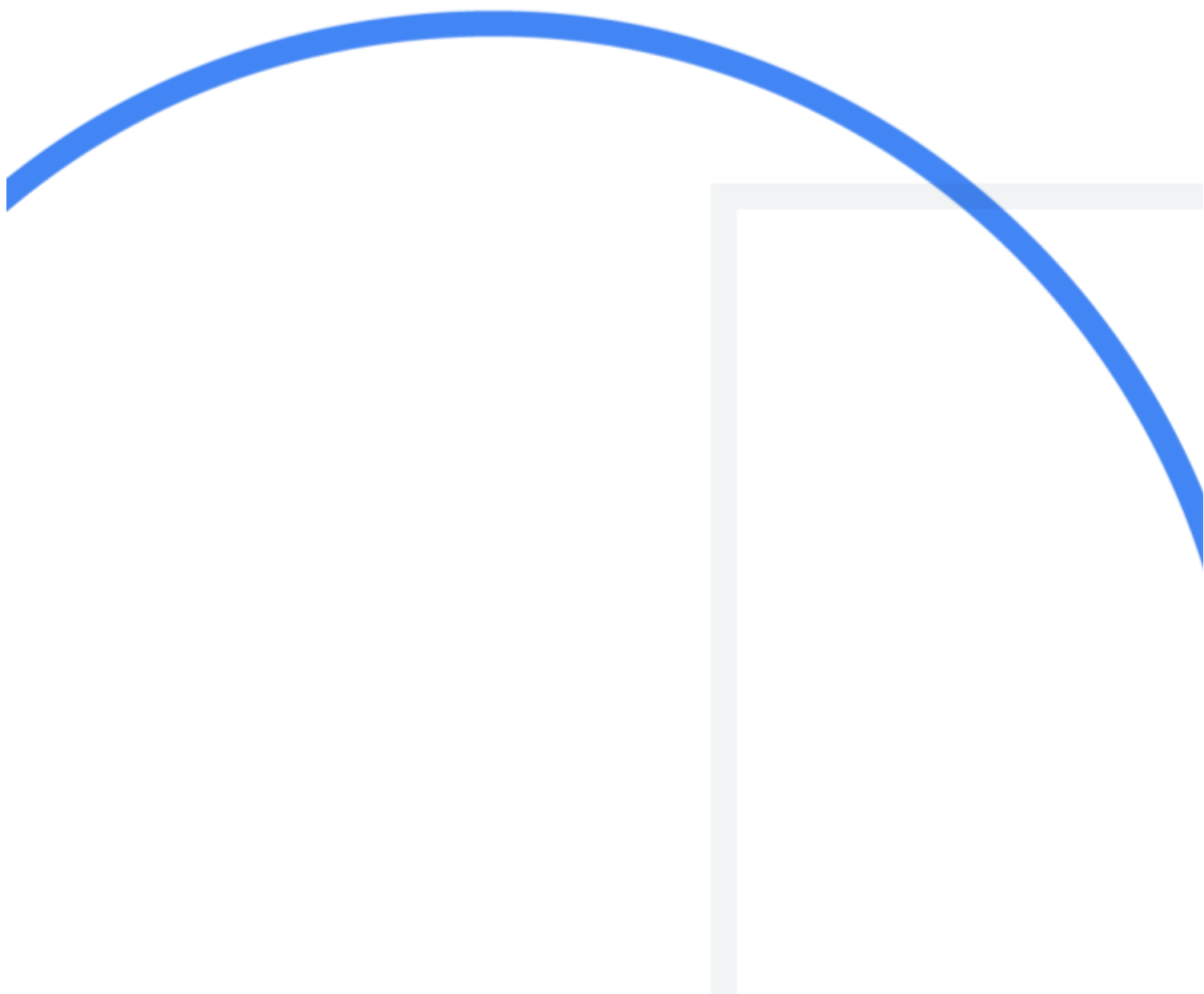


Australia Code of Practice on Disinformation and
Misinformation

Google Initial Report, May 2021



Introduction

Google's mission is to organise the world's information to make it universally accessible and useful. Misinformation and Disinformation run contrary to that mission, across our services. Where our users encounter content that is unreliable or actively designed to mislead them, our products do not serve their purpose of connecting users with content that is relevant and useful to them, and we risk losing their trust. As such, we take these issues immensely seriously. This is particularly true when the content at stake pertains to issues such as health, civics and election, or other issues that may significantly affect the livelihood of our users or the societies that we operate in.

There are inherent tensions that come with fulfilling that mission, especially when it comes to addressing mis- and disinformation. We must strike a careful balance between the free flow of information, safety, efficiency, accuracy, and other competing values.

The product, policy, and enforcement decisions we make in this complex environment are guided by a set of considerations that are consistent across the spectrum of our products and services, and echo a number of the points raised in section 6.1 of this Code:

- **Value openness and accessibility:** We aim to provide access to an open and diverse information ecosystem. But that doesn't mean that anything goes on our services. As we outline below, removal of content is an important lever we use to address information quality. However, it is not the only lever at our disposal, and we use it with caution, particularly in the context of Search. We believe that a healthy and responsible approach to supporting information quality should aim toward keeping content accessible.
- **Respect user choice:** Users who express an intent to explore content that is not illegal or prohibited by our policies should be able to find it, even if all available indicators suggest it is of relatively low quality. We set a higher bar for information quality where users have not clearly expressed what they are looking for.
- **Build for everyone:** Our services are used around the world by users from different cultures, languages, and backgrounds, and at different stages in their lives. Some have always known a world with smartphones, while others have lived most of their lives without access to the web. Our product and policy development, as well as our policy enforcement decisions, take into account the diversity of our users and seek to address their needs appropriately

These have been priorities since our founding. They have guided our evolving approach toward information quality, taking into account shifting user expectations and norms, increasing sophistication of malicious actors, and the evolving nature of the web.

However, there is no single tool or 'silver bullet' that would let us address the whole of the challenges posed by mis- and disinformation: due to their complexities, these issues warrant multiple conjoined efforts within our company (as well as whole-of-society responses involving other industries, civil society, policymakers, and more).

As such, when it comes to addressing these issues across our services, we rely on four complementary categories of efforts (or 'levers') to support information quality and combat mis- and disinformation:

- **“Raise”**: we elevate high-quality content and authoritative sources where it matters most.
- **“Remove”**: we set responsible rules for each of our products and services and take action against content and behaviors that infringe on them.
- **“Reduce”**: We reduce the spread of potentially harmful information where we feature or recommend content.
- **“Reward”**: We set a high standard of quality and reliability for publishers and content creators who would like to monetise or advertise their content.

Each of the products and services we offer has a different purpose, and people have different expectations of what kind of content they will interact with on each. As such, these four levers have different applications across, for instance, Google Search, YouTube, or our Ads Services.

Across these services, while we are cognisant of the differences between misinformation and disinformation as outlined by this Code and by many researchers¹, our product and policy interventions do not typically differentiate between these two harms. That is because the main difference between mis- and disinformation pertains to the *intent to deceive* of the person creating or sharing the content in question – and that is something that we, as a company, can rarely if ever know for certain.

As such, our interventions rather focus on specific *behaviours* or *types of content* that we seek to either prohibit, discourage, or reward – and where there are repeated or egregious violations of our policies, this may lead to taking action against *bad actors* at the level of their accounts or websites².

We aim for these interventions to be as ‘future-proof’ as possible, by which we mean that we design them with the hope that they can be useful across diverse languages, geographies or contexts; and that they take into account the likelihood that bad actors will likely explore new approaches to try and circumvent any intervention we deploy.

In addition to the work that we do on our own services, we work to support the work of others who tackle mis- and disinformation, such as newsrooms, fact-checkers, health-authorities, researchers, and more.

For each of the sections of this report, we will share relevant information about our work across all of these categories, with an eye towards enabling readers to understand the complementary ways with which we tackle these issues. We will primarily focus on Google Search, Ads, and YouTube but will refer to other services where appropriate in light of the specifics of questions.

Importantly, we view our commitments under the Code as a floor, not a ceiling: we fully expect to continue improving our products, policies, enforcement, and transparency work to go beyond the expectations set by this Code, as we adapt to new challenges or identify novel ways to combat mis- and disinformation. This is made possible by the flexibility of this Code, which does not explicitly or implicitly mandate any one specific form of intervention and therefore gives room for signatories to continue experimenting and improving. We also appreciate the opportunity to highlight investments

¹ E.g. [Wardle, Derakhshan, 2017](#)

² For an overview of Actors, Behaviors, and Content vectors in the context of disinformation, see e.g. [Francois, 2019](#)

both in research to further develop our collective understanding of this complex and under-researched issue, and in tangible educational tools such as the [Media Literacy Lab](#) developed by the Alannah and Madeline Foundation (funded by a \$1.4m grant from Google).

Finally, in the interest of simplicity, we note that in the remainder of this document we will refer to “misinformation and disinformation” as “misinformation”. In events where we would like to refer specifically to misinformation as opposed to disinformation (or vice-versa), we will call it out explicitly for readers.

Objective 1: Safeguards against Disinformation and Misinformation

Outcome 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.

As mentioned above, when it comes to addressing the risk of harms that may arise from the propagation of Disinformation and Misinformation on across our services, we rely on four complementary categories of efforts (or 'levers'):

- **“Raise”**: we elevate high-quality content and authoritative sources where it matters most.
- **“Remove”**: we set responsible rules for each of our products and services and take action against content and behaviors that infringe on them.
- **“Reduce”**: We reduce the spread of potentially harmful information where we feature or recommend content.
- **“Reward”**: We set a high standard of quality and reliability for publishers and content creators who would like to monetise or advertise their content.

In the context of YouTube, one of the key success metrics that we apply and that is reported on within the quarterly [YouTube Community Guidelines Enforcement Report](#) is a breakdown of videos that were removed according to how many people had viewed the video before it was removed. In the latest report (spanning October - December 2020), 35.9% of the videos that were removed during that quarter were removed before they had been viewed by a single person worldwide. A further 35.9% of videos were removed with less than 10 views. This data point is an important outcome focused metric that we think demonstrates our efforts to reduce the risk of harm arising from disinformation and misinformation.

Each of the products and services we offer has a different purpose, and people have different expectations of what kind of content they will interact with on each. As such, these four levers have different applications across, for instance, Google Search, YouTube, or our Ads Services and we go into further detail about these product specific measures throughout this report.

Outcome 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.

Across our services, we work to ensure that our policies are publicly available to users and creators in a form that is clear (people can understand what we refer to), predictable (people can determine whether their content or behaviour is likely to violate our rules), and repeatable (our rules should also be generic enough that they can be applied consistently across multiple independent cases around the world).

By way of example – users in Australia and around the world can find our policies on the following webpages:

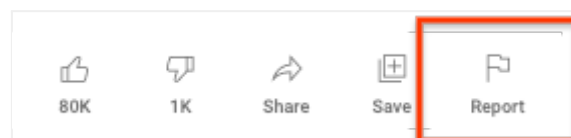
- [Google Ads Policies](#) (applicable to ads showing up on our properties, and the advertisers that run them)
- [Google Publisher Policies](#) (applicable to webpages or videos using our services to monetise traffic, and to their websites or content creators)
- [YouTube Community Guidelines](#) (applicable to YouTube content, users, and content creators)
- [YouTube Monetisation Policies](#) (applicable to YouTube creators who want to monetise their videos via our YouTube Partner Program)
- [Google News Policies](#) (applicable to content and websites in Google News)
- [Google Autocomplete policies](#) (applicable to autocompletions in the Google Search bar)

For more of our policies, see [this page](#).

Outcome 1c: Users can report content or behaviours to Signatories that violates their policies under section 5.10 through publicly available and accessible reporting tools.

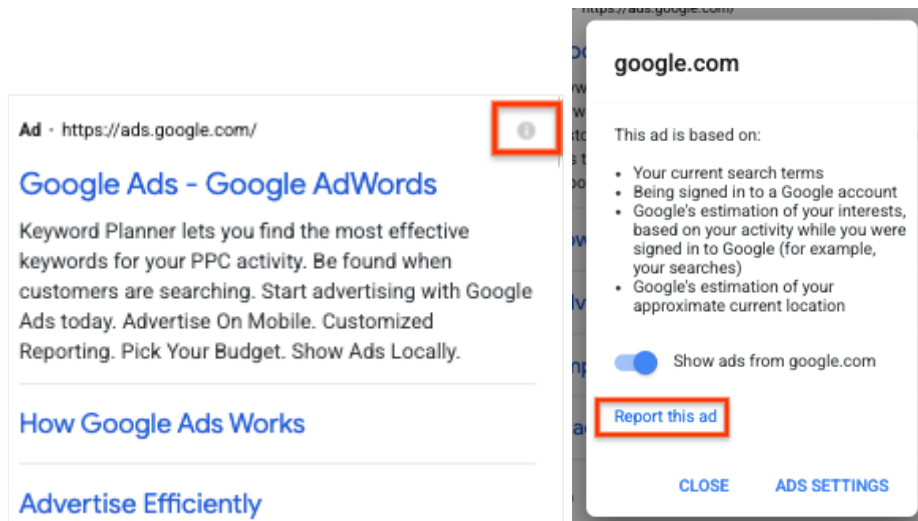
Our products and services enable users to let us know when they believe they have spotted content that violates our policies.

On **YouTube**, users can use our ‘Report’ button to flag issues with any video, including information about which of our policies they believe it is violating.



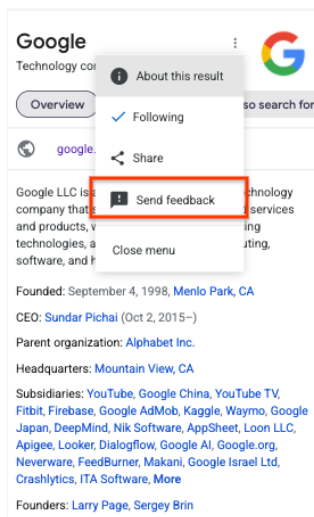
“Report” button on YouTube

On **Google Ads**, users can tap the ‘Why this Ad’ button on the top right corner of an ad and select ‘Report this Ad’ to let us know about ads they believe violate our policies.



Opening the “why this ad” interface and “Report this ad” button

In **Google Search**, we enable users to provide feedback on specific features such as Autocomplete predictions or Knowledge panels; or on a Search results page overall.



“Send Feedback” button in a Knowledge Panel

Outcome 1d: Users will be able to access general information about Signatories’ actions in response to reports made under 5.11.

We regularly report on our work to enforce our content policies across our services:

- Our [YouTube Community Guidelines Enforcement report](#) provides quarterly update on the work we do to enforce our policies on YouTube;
- Our [annual Ads Safety Report](#) provides updates on our work to enforce our policies in Google Ads;
- Our [Google Transparency Report website](#) provides additional information on removals pursuant to valid court or government orders, copyright notices, and more.

- Our Threat Analysis Group Quarterly Bulletin (published on our [Threat Analysis Group blog](#)) provides updates on the actions we take against Coordinated Influence Operations around the world – including but not limited to Australia.

Objective 2: Disrupt advertising and monetisation incentives for disinformation

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

Relevant policies across Google and YouTube

Our [advertising](#) and [monetisation](#) policies prohibit a range of behaviours and types of content that are clearly connected to harmful misinformation, or that commonly overlap with harmful misinformation.

Our policies are put in place for many reasons, including ensuring a good experience for people viewing our ads, preventing user harm, and helping make sure that ads follow applicable laws in the countries where they appear. They also instill trust in our advertiser partners that their ads are running alongside appropriate content. For example, advertisers do not want their ads appearing next to hate speech, misinformation, or other harmful content.

Our policies include:

Google Ads Misrepresentation policy (detailed explanations and examples available [here](#)):

“We want users to trust the ads on our platform, so we strive to ensure ads are clear and honest, and provide the information that users need to make informed decisions. We don’t allow ads or destinations that deceive users by excluding relevant product information or providing misleading information about products, services, or businesses.”

As noted in our help centre, this policy prohibits the following content and behaviours (not a comprehensive list):

- **Coordinated Deceptive Practices:**
 - Coordinating with other sites or accounts and concealing or misrepresenting your identity or other material details about yourself, where your content relates to politics, social issues, or matters of public concern;
 - Directing content about politics, social issues, or matters of public concern to users in a country other than your own, if you misrepresent or conceal your country of origin, or other material details about yourself.
- **Misleading representation:** Making misleading statements, obscuring, or omitting material information about your identity, affiliations, or qualifications
- **Manipulated media:** Manipulating media to deceive, defraud, or mislead others
- **Unreliable claims related to politics, social issues, or matters of public concern:** Making claims that are demonstrably false and could significantly undermine participation or trust in

an electoral or democratic process (e.g. Information about public voting procedures, political candidate eligibility based on age or birthplace, election results, or census participation that contradicts official government records; incorrect claims that a public figure has died, or been involved in an accident)

Google Ads Inappropriate content policy (detailed explanations and examples available [here](#)):

“We value diversity and respect for others, and we strive to avoid offending users, so we don’t allow ads or destinations that display shocking content or promote hatred, intolerance, discrimination, or violence.”

As noted in our help centre, this policy includes the following content and behaviours (not a comprehensive list):

- **Dangerous or derogatory content:**
 - Content that relates to a current, major health crisis and contradicts authoritative scientific consensus
 - Content that incites hatred against, promotes discrimination of, or disparages an individual or group on the basis of their race or ethnic origin, religion, disability, age, nationality, veteran status, sexual orientation, gender, gender identity, or any other characteristic that is associated with systemic discrimination or marginalisation
 - Content that harasses, intimidates, or bullies an individual or group of individuals
 - Content that threatens or advocates for physical or mental harm on oneself or others
- **Sensitive events:**
 - Ads that potentially profit from or exploit a sensitive event with significant social, cultural, or political impact, such as civil emergencies, natural disasters, public health emergencies, terrorism and related activities, conflict, or mass acts of violence

Google Publisher Misrepresentation policy (detailed explanations and examples available [here](#)):

This policy prohibits the following content and behaviours (not a comprehensive list):

- **Dangerous or derogatory content:** we do not allow content that:
 - Incites hatred against, promotes discrimination of, or disparages an individual or group on the basis of their race or ethnic origin, religion, disability, age, nationality, veteran status, sexual orientation, gender, gender identity, or other characteristic that is associated with systemic discrimination or marginalisation.
 - Harasses, intimidates, or bullies an individual or group of individuals.
 - Threatens or advocates for physical or mental harm to oneself or others.
 - Relates to a current, major health crisis and contradicts authoritative, scientific consensus.
- **Misrepresentative content:** we do not allow content that:
 - Misrepresents, misstates, or conceals information about you, your content or the primary purpose of your web destination.
 - Entices users to engage with content under false or unclear pretenses
 - Engages in "phishing" for users' information
 - Promotes content, products, or services using false, dishonest, or deceptive claims
 - Makes claims that are demonstrably false and could significantly undermine participation or trust in an electoral or democratic process

- Falsely implies having an affiliation with, or endorsement by, another individual, organisation, product, or service
- Deceives users through manipulated media related to politics, social issues, or matters of public concern.
- **Coordinated deceptive practices:** we do not allow:
 - Coordinating with other sites or accounts and concealing or misrepresenting your identity or other material details about yourself, where your content relates to politics, social issues or matters of public concern.
 - Directing content about politics, social issues, or matters of public concern to users in a country other than your own, if you misrepresent or conceal your country of origin or other material details about yourself.

Advertiser choice and control:

In addition to these policies, we empower advertisers with choice and control; advertisers can choose whether or not their ads run on certain pages or sites through, for example, keyword, topic or webpage/video exclusions. We are constantly working on improving our offers and to further educate publishers about the tools at their disposal.

This is why in April 2021 we announced the introduction of [dynamic exclusion lists](#), which will help advertisers more easily and efficiently control where their ads show online, and more easily leverage the help and expertise of third party organisations to protect their brands if they choose to do so.

YouTube Partner Program

On YouTube, creators who meet our eligibility criteria can apply to join the [YouTube Partner Program](#), which makes their videos eligible to run advertising and earn money through other sources on the platform. Not only does a channel have to apply, but just because creators are eligible to monetise does not mean that all of their content is necessarily eligible to serve ads and earn money. We have a set of Advertiser-friendly content guidelines and disable advertising from running on videos that violate these policies. Our advertisers are understandably sensitive to what content their ads are placed against. These policies ensure that we strike a balance where advertisers and creators are both able to thrive.

Over the past years, we've taken steps to strengthen our requirements for monetisation so spammers, impersonators, and other bad actors can't hurt our ecosystem or take advantage of good creators. To apply for membership in the YouTube Partner Program, we increased the thresholds needed for channels to be deemed eligible: channels must have generated 4,000 watch hours in the previous 12 months and have more than 1,000 subscribers. After they apply, our review team will ensure their channel has not run afoul of our monetisation, content, and copyright policies prior to admitting them to the program (this includes the Publisher policies mentioned above as well as YouTube's broader community guidelines). As a part of this process, we believe that only creators with sufficient history and demonstrated advertiser safety will receive access to ads and our other monetisation products. In changing these thresholds, we have significantly improved the protections we had in place against impersonating creators.

In addition to these policies, across both Google and YouTube, we provide advertisers with additional controls and help them exclude types of content that, while in compliance with our policies, may not fit their brand or business. These controls let advertisers apply content filters or

exclude certain types of content or terms from their video, display, and search ad campaigns. Advertisers can exclude content such as politics, news, sports, beauty, fashion and many other categories. These categories are listed in our [Google Ads Help Centre](#).

Enforcing our policies to reduce monetisation incentives for misinformation

To make sure that advertisers and publishers on our networks comply with our policies, we continuously monitor our advertising networks for compliance with these policies around the world, using a combination of algorithmic and human reviews.

While bad actors will continue to try to exploit users and Google for their own means, and sites around the world try to monetise harmful or dangerous claims, Google is committed to protecting users, advertisers and publishers from these threats. [In 2020](#), we blocked and removed globally 3.1 billion bad ads - that's more than 5900 bad ads per minute. We also suspended more than 1.7 million advertiser accounts for egregious policy violations. We also blocked or removed over 867 million ads for attempting to evade our detection systems, including cloaking, and an additional 101 million ads for violating our misrepresentation policies. On the publisher side, we removed ads from over 1.3 billion pages that are part of our publisher network (for violating our policies) and took site level action on nearly 1.6 million publishers.

By way of a more specific illustration of the scale of our work in line with the present Code, in March 2021 we blocked over 102,000 creatives from Australian-based advertisers for violating our misrepresentation ads policy. This includes ads violating our misleading representation, clickbait and unacceptable business practices policies.

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

Ever since Google's early days, we have had to deploy rigorous measures to protect the integrity of our services against broad ranges of deceptive practices that could harm our users. These range from webspam (attempts to deceive Search engine ranking systems) to fake engagement on YouTube, fraudulent clicks on Google Ads, artificial traffic on pages seeking to monetise with Google, and more.

While a number of these behaviours are not specific to creators or propagators of misinformation, the Code appropriately notes that vulnerabilities to the integrity of our products and services could be leveraged by these actors and as such merit addressing as part of our response to misinformation.

Our policies prohibit many behaviours aimed at undermining the integrity of our services. Policies that are most relevant to our advertising and monetisation services are already mentioned in response to Objective 2 / Outcome 2 above (e.g. Google Ads misrepresentation or coordinated deceptive practices policy). In this section, we will list other examples across Google Search, News and YouTube – noting that this does not represent a comprehensive list.

Just as for other interventions mentioned in this report, unless stated otherwise, these policies apply to Australian users and content creators and we enforce them diligently, using a mix of automated systems and human reviews.

Google Search: Our [webmaster guidelines](#) prohibit (among others) the following techniques, which could be misused to deceive our ranking systems or abuse our users:

- [Automatically generated content](#)
- Participating in [link schemes](#)
- [Cloaking](#)
- [Sneaky redirects](#)
- [Hidden text or links](#)
- [Doorway pages](#)
- [Scraped content](#)
- Loading pages with [irrelevant keywords](#)

- Creating pages with [malicious behaviour](#), such as phishing or installing viruses, trojans, or other badware
- Abusing [structured data](#) markup
- Sending [automated queries](#) to Google

Where a website violates one or more of these guidelines, Google may take manual action against it. Sites may be submitted for reconsideration once the issue is remedied.

Google News:

- Our [Google News Deceptive Practices policy](#) prohibits sites or accounts that impersonate any person or organisation, or that misrepresent or conceal their ownership or primary purpose. We do not allow sites or accounts that engage in inauthentic or coordinated behaviour that misleads users. This includes, but isn't limited to, sites or accounts that misrepresent or conceal their country of origin or that direct content at users in another country under false premises. This also includes sites or accounts working together in ways that conceal or misrepresent information about their relationships or editorial independence.
- The [Google News Spam and Malware policy](#) prohibits excessive repetitive, duplicate, unoriginal content or artificial freshening, misspellings, grammatical errors, or gimmicky character use. We do not allow links to malware, viruses, or other harmful software. We also do not allow sites engaging in link schemes intended to manipulate a site's ranking in news results or generate unnatural traffic or any other behaviour addressed under our Webmaster guidelines.
- The [Google News Transparency policy](#) notes that visitors to your site want to trust and understand who publishes the content they are consuming, and information about those who have written articles. That's why news sources on Google should provide clear dates and bylines, as well as information about authors, the publication, the publisher, company or network behind it, and contact information.

YouTube: Our [YouTube Community Guidelines](#) include the following policies:

- [Impersonation policy](#): Content intended to impersonate a person or channel is not allowed on YouTube. YouTube also enforces trademark holder rights. When a channel, or content in the channel, causes confusion about the source of goods and services advertised, it may not be allowed.
- [Fake engagement policy](#): YouTube doesn't allow anything that artificially increases the number of views, likes, comments, or other metrics either by using automatic systems or serving up videos to unsuspecting viewers. Also, content that solely exists to incentivise viewers for engagement (views, likes, comments, etc) is prohibited. Content and channels that don't follow this policy may be terminated and removed from YouTube.
- [Spam, deceptive practices, and scam policies](#): YouTube doesn't allow spam, scams, or other deceptive practices that take advantage of the YouTube community. We also don't allow content where the main purpose is to trick others into leaving YouTube for another site. This includes, for instance:
 - Misleading Metadata or Thumbnails: Using the title, thumbnails, description, or tags to trick users into believing the content is something it is not.
 - Comments Spam: Comments where the sole purpose is to gather personal information from viewers, misleadingly drive viewers off YouTube, or perform any of the prohibited behaviours noted above.

- Video Spam, including posting the same content repeatedly across one or more channels or massively posting affiliate content in dedicated accounts.
- Incentivisation Spam: Content that sells engagement metrics such as views, likes, comments, or any other metric on YouTube. This also includes content where the only purpose is to boost subscribers, views, or other metrics (e.g. offering to subscribe to another creator's channel solely in exchange for them subscribing to your channel, also known as "Sub4Sub" content).
- In addition, we [prohibit](#) encouraging the violation of our Terms of Service and posting content previously removed for violating our Terms of Service or content from creators who have been terminated under these Terms.

Across our services – abusive account creation: Our systems rely on a host of inputs about historical use and pattern recognition across various services in an effort to detect if an account creation or login is likely to be abusive and to detect and prevent other types of suspicious conduct. As examples, our systems operate to block “bad” account creation or to close groups of such accounts. We prevent users from creating a large number of Google Accounts in a short time period. If we detect suspicious conduct, we require verification aimed at detecting whether a bot is attempting to access or create an account. There is no panacea, however, for this challenge. We continue to work collaboratively with others to identify inauthentic accounts that are used for improper purposes and deactivate them as soon as possible.

It is worth noting that these policies and other, similar behaviours-focused policies across our services serve as grounds for our actions against Coordinated Influence Operations all around the world – including, but not limited to, foreign and/or state-backed operations.

Our Threat Analysis Group and Trust & Safety teams work to monitor such malicious actors around the globe, disable their accounts, and remove the content that they posted – including but not limited to operations that may affect Australia. We provide monthly updates about these operations in our Threat Analysis Group bulletin ([available on the TAG blog](#)) – for instance, over 2021 thus far:

In January:

- We terminated 4 YouTube channels and 1 advertising account as part of our ongoing investigation into coordinated influence operations linked to Ukraine. This campaign uploaded content in Russian pertaining to current events in Kazakhstan and critical of European Union policies toward Moldova.
- We terminated 5 blogs as part of our investigation into coordinated influence operations linked to Morocco. This campaign uploaded content in Arabic that was critical of the Algerian government. This campaign was consistent with similar findings reported by Facebook.
- We terminated 5 YouTube channels as part of our investigation into coordinated influence operations linked to Brazil. This campaign was linked to a PR firm named AP Exata Intelligence and uploaded content in Portuguese expressing support for several mayoral candidates in Brazil. This campaign was consistent with similar findings reported by Facebook.
- We terminated 6 YouTube channels as part of our investigation into coordinated influence operations linked to Kyrgyzstan. The campaign uploaded content in Kyrgyz critical of the

former President Almazbek Atambayev and the opposition leader Adakhan Madumarov. This campaign was consistent with similar findings reported by Facebook.

- We terminated 3 advertising accounts as part of our investigation into coordinated influence operations linked to Egypt. This campaign was linked to a PR firm named Mubashier and uploaded content in Arabic supportive of the Russian government across several countries in the Middle East.
- We terminated 1 YouTube channel as part of our ongoing investigation into coordinated influence operations linked to Russia. This campaign uploaded content in Russian on current events in Ukraine.
- We terminated 1 YouTube channel, 2 advertising accounts and 1 mobile developer account as part of our ongoing investigation into coordinated influence operations linked to Russia. This campaign uploaded content in Russian on such topics as the U.S. election and the poisoning of Alexei Navalny.
- We terminated 5 YouTube channels as part of our ongoing investigation into coordinated influence operations linked to Russia. This campaign uploaded content in Russian on such topics as the annexation of Crimea and the Syrian civil war.
- We terminated 2 YouTube channels and 1 advertising account as part of our ongoing investigation into coordinated influence operations linked to Russia. This campaign uploaded content in Russian on historical events in Afghanistan, Armenia and Ukraine.
- We terminated 2 YouTube channels as part of our ongoing investigation into coordinated influence operations linked to Russia. This campaign uploaded content in Russian on such topics as the U.S. current events and Alexei Navalny political rallies.
- We terminated 3 YouTube channels as part of our ongoing investigation into coordinated influence operations linked to Iran. This campaign uploaded content in English and was amplifying narratives on regional topics such as Israel, the Nagorno Karabakh conflict, and the war in Yemen. We received leads from FireEye that supported us in this investigation.
- We terminated 2,946 YouTube channels as part of our ongoing investigation into coordinated influence operations linked to China. These channels mostly uploaded spammy content in Chinese about music, entertainment, and lifestyle. A very small subset uploaded content in Chinese and English about the U.S. response to COVID-19 and growing U.S. political divisions. We received leads from Graphika that supported us in this investigation. These findings are consistent with our previous reports in the [Q3](#) and [Q4](#) TAG bulletins.

In February:

- We terminated 5 YouTube channels as part of our ongoing investigation into coordinated influence operations linked to Iran. This campaign uploaded content in English, Farsi, and Bahasa Indonesian on several topics including criticism of Israel and the U.S. election. This campaign was consistent with similar findings reported by Twitter.
- We terminated 5 YouTube channels as part of our ongoing investigation into coordinated influence operations linked to Russia. This campaign uploaded content in Russian specific to narratives around the Russian military. This campaign was consistent with similar findings reported by Twitter.
- We terminated 938 YouTube channels as part of our ongoing investigation into coordinated influence operations linked to China. These channels mostly uploaded spammy content in Chinese about music, entertainment, and lifestyle. A very small subset uploaded content in Chinese and English about the U.S. COVID-19 vaccine rollout and current events. We

received leads from FireEye and Graphika that supported us in this investigation. These findings are consistent with our previous reports in the [Q3](#) and [Q4](#) TAG bulletins.

In March:

- We terminated 13 YouTube channels as part of our investigation into coordinated influence operations linked to Morocco. This campaign uploaded content in Arabic that was supportive of Morocco's government and discussed issues related to regional intelligence agencies as well as Moroccan ownership of the Western Sahara region. This campaign was consistent with similar findings reported by Facebook.
- We terminated 3 YouTube channels as part of our ongoing investigation into coordinated influence operations linked to Russia. This campaign uploaded content in Russian, French, German, English and Spanish about the conflict in Syria, historical footage of the war in Afghanistan, and the civil war in Eastern Ukraine.
- We terminated 34 YouTube channels as part of our ongoing investigation into coordinated influence operations linked to Russia. This campaign uploaded content in Russian that was critical of the protests in support of Alexei Navalny.
- We terminated 33 YouTube channels and 2 advertising accounts as part of our ongoing investigation into coordinated influence operations linked to Myanmar. This campaign uploaded content in Burmese about the military coup in Myanmar.
- We terminated 682 YouTube channels and 2 advertising accounts as part of our ongoing investigation into coordinated influence operations linked to China. These channels mostly uploaded spammy content in Chinese about music, entertainment, and lifestyle. A very small subset uploaded content in Chinese and English about China's COVID-19 vaccine efforts and social issues in the U.S. These findings are consistent with our previous reports in the [Q3](#) and [Q4](#) TAG bulletins.

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.

Part of our approach to this goal lies in the very design of our ranking systems. Every time a user searches for something on Google, our algorithms work to understand the meaning of their query and to rank the pages in our index according to signals including their relevance to the user's expectation (does the webpage content cover the subject matter the user is asking about? Are the user's keywords present in this webpage?) and their authoritativeness (Google's estimate of the expertise, authority, and trustworthiness of a web page, based on factors like whether users and other sites on the web seems to trust this web page³). We continuously improve upon these ranking systems – for instance, in 2019 only, we made more than 3,600 improvements to our search algorithms.

For subject matters related to news and civic issues, we realise that our users have high expectations when it comes to the trustworthiness of the results we present them. We give more importance to factors related to authoritativeness in response to such queries. This, as well as all other goals and principles of Google Search, is outlined in our [Search Quality Raters Guidelines](#) – which also inform our approach to ranking content on Youtube.

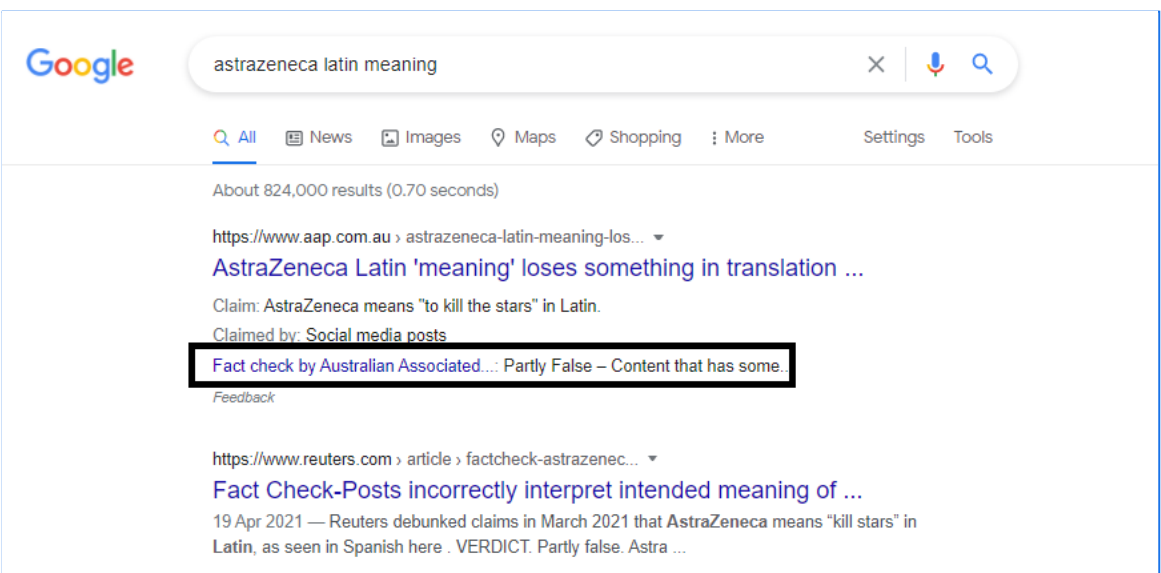
Our ranking systems are also designed to reduce the spread of borderline content—content that comes close to violating our policies but doesn't cross the line. In particular, on YouTube, In 2019, we [announced](#) changes to our recommendation systems to reduce the spread of borderline content, resulting in a 70 percent drop in watchtime on non-subscribed, recommended content in the US that year. We saw a drop in watchtime of borderline content coming from recommendations in other markets as well. While algorithmic changes take time to ramp up and you might see consumption of borderline content go up and down, our goal is to have views of non subscribed, recommended borderline content below 0.5%. We have now rolled out changes to our recommendation system to reduce borderline content in every market where we operate, including Australia. We continue to fine tune our systems globally to keep pace with the fluid nature of problematic content on YouTube.

³ For more information on how we work to deliver reliable information in Google Search, see <https://blog.google/products/search/how-google-delivers-reliable-information-search/> or g.co/howsearchworks

In addition, the design of our service typically presents users with a wide array of webpages and information to choose from at any given point of their navigation on Google or YouTube. This gives users access to a plurality of perspectives, helping them get the full picture on content that is of interest to them.

Furthermore, we have developed many features and interventions dedicated to providing more context to users and making sure that they we do our utmost to elevate authoritative sources in response to searches or browsing intents that are related to topics such as health, civic participation, current events, and other issues where users want content that they can trust. Below, we list examples of these features and interventions – all of which are available to Australian users of our services.

- **Surfacing Fact-Checks on Google Search:** in 2016, we launched the Fact Check label in Google News and Search to make it easier for people to find articles that fact check public information, ranging from claims to public statements to statistics. It is currently used by many fact-checking organisations around the world, including in Australia. To discover Fact-Checks on Google, users merely have to browse Google Search or News and look for links that show a “fact-check” tag or snippet. This feature is open to all publishers who signal their use of fact-checking by using a dedicated html mark-up and comply with our policies on fact-checking. Over time, we have made it even easier for fact-checkers to mark-up their content by creating a [dedicated API](#), which provides an easier alternative for publishers of fact-checks than adding HTML mark-ups. It is possible to explore the fact-checks that we index by visiting <https://toolbox.google.com/factcheck/explorer>.



Screenshot: Fact-check in Search results on Google in Australia

- **Elevating original reporting in Google Search:** over the past years, we have made changes to our search rater guidelines to help us better recognise original reporting, surface it more prominently in Search, and ensure it stays there longer. This means readers interested in the latest news can find the story that started it all, which can provide them

with important context to understand the news—and means news publishers can benefit from having their original reporting more widely seen.

- **Full Coverage in Google News:** to help users access context and diverse perspectives about the news stories they read, the “Full Coverage” feature in Google News lets users explore articles and videos from a variety of publishers related to an article or news story of their choice. The “Full Coverage” feature is not personalised, and accessible in one click or tap from most articles in Google News’ “For You” and “Headlines” tabs.
- **Information panels on Google Search and YouTube:** Google and YouTube’s information panels help users find context about topics related to their searches or to the videos that they are watching.
 - On Google, [Knowledge panels](#) are automatically generated information boxes that appear in response to searches for entities (people, places, organisations, things) that are in our Knowledge Graph. They are meant to help users get a quick snapshot of information on a topic based on Google’s understanding of available content on the web.
 - On YouTube, [Information panels giving topical context](#) may appear on top of search results or under a video when users search or watch videos related to topics prone to misinformation, such as the moon landing or climate change, so as to provide more context on that topic from independent, third-party partners. These information panels show regardless of what opinions or perspectives are expressed in a video.
- **Additional ranking safeguards during breaking news events on Google and YouTube:** During breaking news or crisis situations, stemming the tide of misinformation can be challenging. Speculation can outrun facts as legitimate news outlets on the ground are still investigating. At the same time, malicious actors are publishing content on forums and social media with the intent to mislead and capture people’s attention as they rush to find early information online. Over the years, we have worked to improve our ranking systems across Google and YouTube so that they detect such breaking news contexts and optimise for elevating authoritative sources, rather than risk showing more recent but less reliable content.
- **Breaking News and Top News Shelves on YouTube:** On YouTube, we have worked to launch several product features that prominently surface news content from only authoritative sources. Two such cornerstone products have been the Top News shelf in YouTube search results and the Breaking News shelf on the YouTube homepage. The Top News shelf triggers in response to certain search queries that have news-seeking intent (see screen shot below). Our systems rely on a variety of signals that we derive from Google News and from our own systems when a user might be seeking information on news-related topics. Given the user’s intent is to find out more information in this domain where the authority of content and source is particularly important, we display a clearly delineated shelf in the product where we only surface content from authoritative, verified news channels. The Breaking News shelf triggers when there is a significant news event happening in a specific country. This shelf triggers automatically and is shown prominently on the YouTube homepage to all users in the country. Similar to the Top News shelf, only content from authoritative sources is eligible to be surfaced in the Breaking News shelf.

Case study: Addressing COVID-19 Misinformation

Since the beginning of 2020 and in response to the COVID-19 pandemic, we have deployed many of the resources outlined above in this section and in others to help protect the integrity of our services, elevate authoritative information and context to our users, and remove harmful medical misinformation from our services – in Australia and abroad. In this case study, we will provide a concrete overview of the actions that we took.

Elevating authoritative content:

Since the beginning of the pandemic, search interest in COVID-19 has grown in Australia and around the world. In response, we have worked to help people find the information they need across Google and YouTube – including by partnering with health organisations and governments to bring our users authoritative information in a rapidly changing environment.

- In **Google Search**, we have introduced a comprehensive experience for COVID-19 that provides easy access to information from health authorities alongside new data and visualisations. This new format organises the search results page to help people easily navigate resources and makes it possible to add more information as it becomes available over time. This experience comes as a complement to pre-existing work on Google Search and Google News, mentioned above in this report, to recognise sensitive events and contexts, and our systems are designed to elevate authoritative sources for those classes of queries. In December 2020, we [announced](#) that beginning in the United Kingdom, we're launching a new feature on Search so when people look up information for COVID-19 vaccines, we will surface a list of authorised vaccines in their location, as well as information on each approved vaccine. As other health authorities begin authorising vaccines, we'll introduce this new feature in more countries – including Australia.
- On the **Google HomePage**, in partnership with the World Health Organization and other health authorities, including the Australian Department of Health, we have promoted important guidance to prevent the spread of COVID-19. The efforts, including prevention tips and messaging on our homepage, have launched in more than 100 countries to date, including Australia.
- Across **YouTube**, we are elevating authoritative sources such as the WHO and the Australian Department of Health to help users get the latest COVID-19 information. We display information panels linking to global and locally relevant health officials on our homepage, and in panels that appear on videos and in search results about COVID-19. With countries approving vaccines and countries enacting their vaccination plans, we are adding authoritative vaccination information from local health authorities to these info panels, as it is made available to the public. Collectively, our panels have served over 400 billion impressions around the world. We have also donated ad inventory to governments and NGOs to help give their public health messages about COVID-19 more visibility on YouTube. In addition, YouTube elevates content from authoritative channels such as news organisations or health authorities when our systems detect that a user's search is health-related. Finally, to engage the vibrant YouTube community and to support global leaders in their urging people to stay home and help flatten the curve of COVID-19, since March 2020, YouTube partnered with creators around the world to launch our #StayHome #WithMe campaign. YouTube engaged hundreds of creators and influencers across the globe to film public service announcements to get young people to stay home and work out, study, learn, craft, and listen to music with their favorite YouTube creators virtually. In Australia, we worked with popular creators like Chloe Ting (19.7m subscribers) and WAFL player Tayla Harris.

- In **Google News**, we have created a new COVID-19 section with links to up-to-date, relevant stories from the international to local levels from a variety of authoritative sources. The section is now available to users across 40 top impacted markets, including Australia, and puts local news front and centre by highlighting stories about the virus from local publishers in the reader's area.
- A **website**, which provides resources dedicated to COVID-19 education and prevention, has also been released. It is available on www.google.com/COVID-19 in more than fifty countries. It is currently available in Australia.
- Since the beginning of the pandemic, we've helped more than 100 government agencies and global non-governmental organisations run critical public service health announcements through our \$800 million **Ad Grants Crisis Relief program**. Grantees will be able to use these funds throughout 2021, including for vaccine education and outreach campaigns. In January 2021 we announced an additional \$100 million in Ad Grants for the World Health Organization and nonprofits like Australian Mercy, the YWCA, MedicAlert and Wesley Medical Research. The Australian Federal Government and Department of Health have thus far received \$4.8 million worth of Ad Grants. Resulting from the program, as of March 2021, these Australian governments and authorities have generated 20.6 million impressions and 2.3 million clicks, providing important health information across Australia. In addition, we've given a total of \$48m to Australia-based nonprofits from March 2020 - March 2021, many of whom are serving populations in increased need due to the crisis.

Detecting and removing COVID-19 misinformation

In addition to elevating authoritative information, we take active steps to detect and remove COVID-19 related misinformation that contradicts guidance from health authorities and may result in real-world harm.

- On **YouTube**, our Community Guidelines prohibit content that encourages dangerous or illegal activities that risk serious physical harm or death, including certain types of medical misinformation. As the COVID-19 situation has evolved, we have partnered closely with global and local health authorities to ensure our policy definition and enforcement is effective in removing violative content where there is a serious risk of egregious harm. This work has evolved into a comprehensive [COVID-19 medical misinformation policy](#), which prohibits, for example, content that denies the existence of the coronavirus or encourages the use of home remedies in place of medical treatment. We also prohibit content that explicitly disputes the efficacy of global or local health authority advice regarding social distancing that may lead people to act against that guidance. On October 14, we further expanded this policy to include a set of claims relating to a potential vaccine that contradict expert consensus from local health authorities or the WHO. For example, we will remove content with claims that the COVID-19 vaccine will kill people who receive it. In addition, we continue to improve upon the work mentioned above in this report to [reduce recommendations](#) of borderline content—content that comes close to violating our policies but doesn't cross the line—such as certain types of potentially harmful misinformation.
- On **Google Search**, our [medical topics policy](#) applies to information we've highlighted via our COVID-related Search features. We strive to show information that reflects scientific consensus and evidence-based best practices, since we consider this content high quality. To this end, if this

highlighted information runs contrary to general scientific consensus, we reserve the right to correct or remove the information from the feature.

- On our **Google Ads and Monetised services (AdSense)**, our harmful health claims policies for both [publishers](#) and [advertisers](#) prohibit content about a current, major health crisis that contradicts authoritative, scientific consensus. This builds upon pre-existing policies, such as our AdSense [Dangerous or Derogatory Content policy](#) which prohibits monetisation of content "*promoting or advocating for harmful health or medical claims or practices.*" Under these policies, we demonetise webpages or videos that include claims about the propagation of COVID-19 that contradict the WHO guidance, such as theories involving 5G towers as a transmission vector.
- On **Google Ads and Google Shopping**, our Sensitive Events [policies](#) (also mentioned above in this report) do not allow ads that potentially profit from or exploit a sensitive event with significant social, cultural, or political impact, such as civil emergencies, natural disasters, public health emergencies, terrorism and related activities, conflict, or mass acts of violence—and we have begun treating the COVID-19 crisis as a sensitive event all around the world (including in Australia) by the end of January 2020. Over time, we started phasing in allowances for COVID-related ads from government organisations, healthcare providers, non-governmental organisations, intergovernmental organisations, verified election ads advertisers and managed private sector accounts with a history of policy compliance who want to get relevant information out to the public. Ads that are allowed still have to abide by the policies mentioned above in this case study and in the rest of this report.
- Under our **Google Ads policies**, in March 2021, we've blocked or removed over 11,000 coronavirus-related ads from Australia-based advertisers for policy violations including price-gouging, capitalising on global medical supply shortages, making misleading claims about cures.

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. Signatories detail policies that provide users with information about the source of Political Advertising carried on digital platforms.

In November 2020, in order to bring greater transparency to political advertising and to support the integrity of Australian elections, we launched [new policies](#) for Australian election ads (ads that feature a political party, current elected officeholder, or candidate for the Federal House of Representatives or Senate).

To enforce these policies, we launched additional identity verification requirements (outlined below) for advertisers who would purchase election ads on Google in Australia; in-ad “paid for by” disclosures to inform users of the source of election ads that they see on our services; as well as a [Transparency Report and searchable ads library](#) specifically focused on Australian election ads.

Verification process for Australian advertisers:

Prospective advertisers - falling under the categories as outlined below - who desire to run election ads in Australia are required to go through a verification process which takes place as follows:

Advertisers are required to provide the following information to verify their eligibility to run election ads:

- Google Ads Customer ID
- Email address
- Name of advertiser
- Registration number: For example, Australian Company Number or Australian Business Number (if applicable)
- Proof of organisation: advertisers must submit:
 - A certificate of incorporation or registration, GST registration letter from the Australian Taxation Office or party extract from the Australian Electoral Commission if they are a company, partnership, registered Australian body, association, co-operative, non-profit organisation or registered political party.
 - A certificate or document from a public, government, or judicial authority, such as a letter of approval from an official body; or a document establishing the

organisation's name and address, such as a bank/credit card statement or a utility bill if they are any other organisation type not listed above.

- Official address (must be an Australian address)
- Attestation that the authorised representative is a citizen or permanent resident of Australia.
- Attestation that the organisation applying for verification is registered or based in Australia and legally permitted to run election ads in Australia
- Attestation that the organisation applying for verification has or will satisfy the appropriate disclaimer requirements under Australian law
- Agreement to the Google Election Advertising Terms.

Once Google verifies the advertiser's eligibility to run election ads, they receive an email and an in-account notification (in a red bar at the top of the Google Ads account), that guides them through the identity verification process in Google Ads. Verifying their identity can take up to 5 business days.

To verify, an authorised representative of the organisation will be asked for the following:

- Organisation name
- Organisation address (must be an Australian address)
- Proof of organisation: the representative must submit
 - A certificate of incorporation or registration, GST registration letter from the Australian Taxation Office or party extract from the Australian Electoral Commission if they are a company, partnership, registered Australian body, association, co-operative, non-profit organisation or registered political party.
 - A certificate or document from a public, government, or judicial authority, such as a letter of approval from an official body; or a document establishing their organisation's name and address, such as a bank/credit card statement or a utility bill if they are any other organisation type not listed above.
- Authorised representative name (this must exactly match the name on the government-issued photo ID they submit)
- A copy of the authorised representative's government-issued photo ID: A current Australian passport, Australian driver's license or Australian photo card
- Attestation that the authorised representative is a citizen or permanent resident of Australia.
- Attestation that the organisation applying for verification is registered or based in Australia and legally permitted to run election ads in Australia
- Attestation that the organisation applying for verification has or will satisfy the appropriate disclaimer requirements under Australian law
- Agreement to the Google Election Advertising Terms.

Our teams are trained to handle this process at scale across Australia, and are equipped to respond to related questions from the political parties and candidates participating in, and institutions responsible for, Australian elections.

As of mid-April 2021, 30 advertiser accounts have been successfully verified for the purpose of running [election ads](#) in Australia.

It is important to note that Google's broader ads policies, as described above, also apply to election ads. For instance the prohibition on ads or destinations that seek to deceive users or provide misleading information also applies to all ads, including but not limited to election ads or ads that

deal with political topics. In addition, in order to improve trust around election ads and digital advertising, we only allow [election ads targeting](#) based on general geographic location (down to postal code), age, gender, and context. We also do not allow personalised ads targeting by [political affiliation](#).

Further to our commitment to increase transparency, choice and control for our users, in 2020, we began to introduce a new advertiser verification program to capture more information from all of our advertisers about their identities and businesses. We verify identification and then disclose the advertiser behind specific ads in the 'Why this Ad' (in some places, "About this Ad") menu.

- This verification and disclosure feature applies to the ads that Google shows on our own properties and those of our publishing partners.
- The program is being phased in on a rolling basis geographically, and began rolling out in Australia in the first quarter of 2021.
- We believe giving people the ability to see who the advertiser responsible for specific ads not only helps users make more informed decisions but is also an important component of transparency in the broader ecosystem.

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 have a keen commitment to supporting efforts to deepen all of our collective understanding of disinformation and misinformation specifically and have made considerable efforts to expand knowledge about public health issues during the COVID-19 pandemic through our collaboration with authoritative global (e.g the World Health Organization) and local sources (Federal and State Departments of Health). Examples of these efforts include;

- The provision of funding to the [Australian Science Media Centre](#), creators of [Scimex.org](#). They worked with more than 85 COVID-19 experts and organised 60 [expert reactions](#) from leaders in the fields of science and health, explainers, briefings and myth busters on COVID-19, all of which informed more than 30,000 news stories in Australia and overseas. Google funding also enabled the Centre to hire an additional media officer for six months.
- With support from the Google News Initiative, [First Draft](#) launched its first bureau in APAC based at the Centre for Media Transition at the University of Technology Sydney.
- We worked with the ABC to launch a dedicated News Briefing on the Google Assistant focused on COVID-19.
- Alongside the Museum of Australian Democracy, we funded [research](#) undertaken by Western Sydney University and the Queensland University of Technology seeking to identify the knowledge and skills needed to advance the media literacy of young Australians.
- We provided \$1.4m of funding to the Alannah and Madeline Foundation to develop the [Media Literacy Lab](#); an interactive learning platform that empowers young Australians to think critically, create responsibly, be effective voices and active citizens online.
- In 2020, Google funded the provision of School Grants through Project Rokit which enabled them to deliver a series of interactive online safety webinars to support and impact schools during the period of remote learning due to the COVID-19 pandemic. Over 4,600 students, most of whom were from rural and regional communities, benefited from this valuable guidance. This program built upon an earlier investment in Project Rokit to produce two series of [Project Rokit TV](#); a series of short form videos exploring delicate topics such as different communication styles, online reputation management, the risks and rewards of 'standing up' and supporting friends who are the subjects of rumours.
- Back in 2016, Google provided seed funding of \$1.4m to the Alannah and Madeline

Foundation to develop the [eSmart Digital Licence](#), Australia's number 1 selling online safety education program. Now with three separate curricula for junior primary, primary and secondary school aged children, topics include communicating safely online, searching and researching, creating and sharing, discerning real vs fake content and more. More than 260,000 Australian students are enrolled in this program.

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

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

The present report constitutes our initial filing to report on the efforts we have undertaken to address misinformation in Australia.

We will add more content under this objective in future reports.

Concluding remarks

Over the course of this report, we have described the work that we do to address the risk of harms arising from misinformation and disinformation across our services using our raise, reduce, remove, and rewards levers and provided both examples and quantitative data points that outline how this work applies in practice, in Australia and beyond. We have also outlined how we work with others to support their own work, as misinformation and disinformation are whole-of-society challenges that we need to address together with civil society, newsrooms, academia, government, and more.

We are cognisant that our job is not over and will continue to invest in developing more ways to protect our users and the integrity of our services, while preserving diverse expression and debate online. As such, we welcome the Australian Code of Practice's framing of the main threats and equities that we and others in industry should focus on, and look forward to continuing to report on our work as part of our commitments.