

Australian Code of Practice on Disinformation and Misinformation
Microsoft and LinkedIn
Annual Transparency Report
May 2022



Summary

Microsoft is pleased to file this report on our commitments under the Australian Voluntary Code of Practice on Disinformation and Misinformation.

Microsoft's mission is to empower every person and every organisation on the planet to achieve more. We offer an array of services, including cloud-based solutions that provide customers with software, services, platforms, and content, and we provide solution support and consulting services. We also deliver relevant online advertising to a global audience.

Our products include operating systems, cross-device productivity applications, server applications, business solution applications, desktop and server management tools, software development tools, and video games. We also design and sell devices, including PCs, tablets, gaming and entertainment consoles, other intelligent devices, and related accessories.

At Microsoft, we are committed to instilling trust and security across our products and services, and across the broader web. We also recognise that fighting disinformation is a key element to creating a trustworthy and safe online environment. Microsoft has a long history of working closely with governments, industry, civil society organisations, academia, and other stakeholders to ensure the integrity and security of our services and the online space more generally. Our efforts are not limited to disinformation, however; instead, they extend to many other related areas designed to promote the integrity, trust, security, and safety of our online services.

As the Code states, "...the types of user behaviours, content and harms that this code seeks to address will vary greatly in incidence and impact amongst the diverse range of services and products offered by different digital platforms." In Microsoft's case, the majority of our services are used by enterprise customers or by individuals acting in a professional capacity.

Disinformation is not typically spread through business-to-business interactions, or between individuals interacting in a professional context. Furthermore, because bad actors typically use online services directed at consumers to disseminate disinformation, business-oriented online services are less likely to experience disinformation on their services.

Bearing in mind these considerations, the remainder of this report focuses on how Microsoft is working as a company and across our relevant services to implement the commitments in the Code and to combat disinformation online.

The Microsoft services in scope of the Voluntary Code are:

- **Microsoft Bing:** a web search engine which provides a variety of services including web, video, image, and map search products. Microsoft Bing does not host the content appearing in search results, does not control the operation or design of the indexed websites and has no ability to control what indexed websites publish

¹ Australian Code of Practice on Disinformation and Misinformation, Preamble, p. 3.

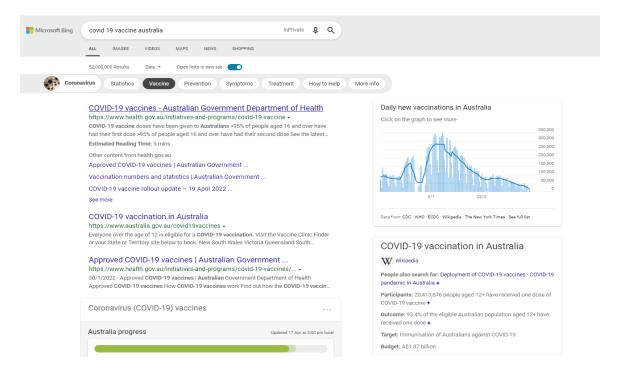


- Microsoft Advertising: Microsoft's proprietary online advertising network, which
 serves the vast majority of ads displayed on Microsoft Bing, and provides advertising
 to most other Microsoft services that display ads, as well as many third-party services
- **Microsoft Start:** a service which delivers high quality news across web and mobile for Microsoft customers and syndication partners
- **LinkedIn:** a real identity professional network where members can find jobs, connect, and strengthen professional relationships, and learn the skills they need to succeed in their careers. It operates via websites and mobile apps and includes user-generated content

Disinformation comes in many forms, and no single technology will solve the challenge of helping people decipher what is true and accurate. We manage risk of disinformation and misinformation on our products through our policies and practices, our innovations in technology, and our research and partnerships.

Unless otherwise specified, data provided is for 2021 calendar year.

During the reporting period, Microsoft has taken action to reduce the impact of disinformation and misinformation on **Microsoft Bing** such as through continued improvement of our ranking algorithms to ensure that the most authoritative relevant content is returned at the top of search results. Microsoft Bing also continues to work towards providing users with authoritative content using enhanced search results such as answer boxes and similar experiences. As an example, below is a screenshot that shows right-rail rich results and additional relevant content interspersed throughout the search results page.





On **Microsoft Start**, we have introduced policies to specifically address disinformation and misinformation on three clear and well-defined misinformation narratives, namely: COVID-19 (introduced March 2021), QAnon (introduced May 2021), and Russia-Ukraine (introduced February 2022). These first three topics were chosen based on the potential for real-world harm and the perniciousness of their spread across the globe. Microsoft Start has also addressed violations to our community guidelines through takedowns.

In 2021, **Microsoft Advertising** took down more than 3 billion ads globally for various policy violations, almost twice as many as in 2020. We introduced <u>advertiser identity</u> <u>verification</u> in seven markets, including Australia, to ensure customers see ads from trusted sources by requiring selected advertisers to establish their identity as a business or as an individual.

From January to June 2021, **LinkedIn** globally blocked more than 15 million fake accounts and removed more than 147,000 pieces of misinformation. Over the same period, LinkedIn blocked approximately 120,000 fake accounts attributed to Australia and removed approximately 2,000 pieces of misinformation reported, posted, or shared by Australian members.

Microsoft and LinkedIn are signatories to the Statement of Intent on the Election Working Arrangements between the Electoral Australian Commission and Online Platforms. Microsoft and LinkedIn worked closely with the Australian Electoral Commission and established dedicated arrangements to manage content referrals related to foreign disinformation and breaches of the Electoral Act during the 2022 Federal Election campaign period. Neither Microsoft Advertising nor LinkedIn accept political ads.

As published on our <u>Microsoft On the Issues blog</u>, in response to the invasion of Ukraine in 2022, we swiftly moved to reduce the exposure of Russian state propaganda, as well to ensure our own platforms do not inadvertently fund these operations. This has included:

- Not displaying any Russia Today (RT) and Sputnik content in Microsoft Start
- Removing RT news apps from our Windows app store
- Further de-ranking RT and Sputnik news sites' search results on Microsoft Bing so that it will only return links when a user clearly intends to navigate to those pages
- Banning all advertisements from RT and Sputnik across our ad network and not placing any ads from our ad network on these sites

These actions were taken across the Australian versions of our services and services available to Australians.

Commitments under the Code

1a: Contribute to reducing risk of harm by	Microsoft Bing, Microsoft Start, Microsoft
adopting scalable measures	Advertising, LinkedIn



1b: Users informed about types of behaviours and content prohibited/managed	Microsoft Start, Microsoft Advertising, LinkedIn
1c: Users can report content that violates policy through accessible reporting tools	Microsoft Bing, Microsoft Start, Microsoft Advertising, LinkedIn
1d: Users can access general information about response	Microsoft Bing, Microsoft Start, Microsoft Advertising, LinkedIn
2: Advertising and/or monetisation incentives reduced	Microsoft Advertising, LinkedIn
3: Risk of inauthentic behaviours undermining integrity and security of services/products reduced	Microsoft Bing, Microsoft Advertising, LinkedIn
4: Users more enabled to make informed choices about sources of news and factual content and to identify misinformation	Microsoft Bing, Microsoft Start, LinkedIn
5: Users better informed about source of Political Advertising	Microsoft Advertising, LinkedIn
6: Support efforts of independent research	Microsoft
7: Public access to measures to combat disinformation and misinformation	Microsoft Bing, Microsoft Start, Microsoft Advertising, LinkedIn

Reporting Against Commitments

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.

Microsoft reduces the risk of harms that may arise from the propagation of disinformation and misinformation on **Microsoft Bing, Microsoft Start, Microsoft Advertising** and **LinkedIn** through the application of our internal policies and scalable measures.

Microsoft Bing

As an online search engine, the primary objective of **Microsoft Bing** is to connect users with the <u>most relevant search results</u> from the web - providing easy access to content produced by and hosted by third party publishers.

As an algorithmically driven service, Microsoft Bing doesn't host the content or control the operation or design of the indexed websites and has no control over what indexed websites publish. As long as the website continues to make the information available on the web and to crawlers, the information will be generally available through Microsoft Bing or other search engines.



The section "How Bing Ranks Your Content" in the Microsoft Bing Webmaster Guidelines details the main parameters Microsoft Bing uses to rank content, prioritising relevance and quality and credibility, which leads to the demotion of low-authority content. The Webmaster Guidelines also detail the types of inauthentic behaviours (spam) that may lead to additional ranking penalties

Disinformation may at times appear in both organic and paid search results, and we take active steps to counter it.

It is worth emphasising, however, that addressing disinformation in organic search results often requires a different approach than may be appropriate for other types of online services, such as social media services.

- Importantly, blocking content in organic search results based solely on the truth or
 falsity of the content can raise significant fundamental rights concerns relating to
 freedom of expression and the freedom to receive and impart information
- For example, even with known fake articles, such as "Islamist Militants Set Fire to Notre Dame," a researcher or journalist may actually want to find such unreliable information for legitimate reasons

So, while Microsoft Bing generally strives to rank its organic search results so that trusted, authoritative news and information appear first, and provides tools that help Microsoft Bing users evaluate the trustworthiness of certain sites, we also believe that enabling users to find all types of information through a search engine can provide important public benefits.

In response to the COVID-19 pandemic, Microsoft Bing introduced new methods to promote reliable content for Australian users.

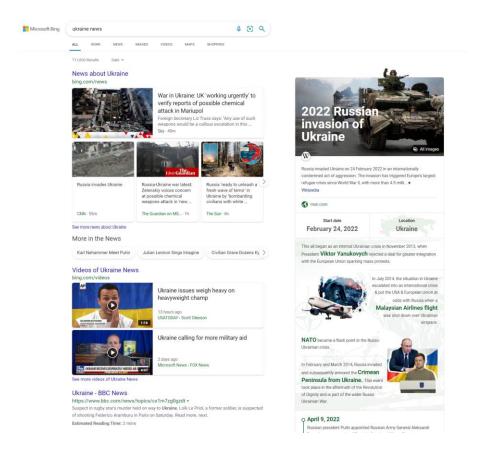
- Reliable health information is provided through public service announcements for certain COVID-19 queries
- Reliable information can also be found near the top of search results pages and in sidebar windows
- Prioritising trusted news sources and piloting algorithmic defences to help promote reliable information

In response to Russia's invasion of Ukraine in 2022, Microsoft Bing is monitoring closely and working to promote authoritative content related to the situation.

- Microsoft Bing has taken steps to algorithmically boost authority signals that as at mid-May have downgraded less authoritative information for over 230,000 distinct queries impacting over 31 million impressions in all languages
- This has helped ensure that Microsoft Bing is promoting authoritative news sources, timelines, and other factual information at the top of algorithmic search results

A screenshot illustrating how Microsoft Bing promotes relevant news content is below. Microsoft Bing has been working internally and with third party experts to identify new threats in this space in order to quickly address any ranking issues that are found.





Microsoft Start

Microsoft Start delivers high-quality news across web and mobile experiences for Microsoft as well as a growing number of syndication partners. Microsoft Start's model reduces risk of disinformation and misinformation being propagated. Misinformation in our licenced content feed has been exceedingly rare.

- Our content providers are vetted and are contractually committed to a high editorial standard and must adhere to a strict set of standards that prohibit false information, propaganda and deliberate misinformation
- Microsoft Start is free to download, with no limits on number of articles or videos a user can view
- We have specific policies for managing misinformation relating to three clear and well-defined misinformation narratives with potential for real-world harm - COVID-19, Russia-Ukraine and QAnon - which includes disabling comments for certain articles to reduce propagation of disinformation
 - In response to the invasion of Ukraine in 2022, we quickly developed and published a misinformation trait – which defines what can and cannot be said about this topic - to prevent the Microsoft Start Community from becoming a platform for disinformation
 - For instance, we forbid claims that the Ukrainian government are Nazis and paedophiles



In the few occasions globally (fewer than ten) that misinformation has been flagged in articles during and prior to the reporting period, Microsoft has responded with reactive takedowns. There have been no instances in Australia.

Microsoft Start Community was introduced in September 2020 and launched in Australia in May 2021. It supports diverse, authentic conversations and content about issues and events. This is a safe, inclusive, and respectful forum where participants are responsible for their posts and how they treat one another. Our <u>Community Guidelines</u> are designed to uphold these values and we strive to provide transparency and clear guidance on how to comply with them.

- If a contribution is flagged, it will be reviewed. If it does not meet the community guidelines it will be removed
- User activity feed shows if any comments have been removed and users are able to appeal the decision
- The Comments function was relatively nascent during the reporting period

When necessary, Microsoft Start will suspend a user's ability to comment. Continued refusal to meet standards may result in permanent ban, which users have an opportunity to appeal. A misinformation trait will define what can and cannot be said on our platform about a particular topic. We rely on recognised independent authorities to identify and verify misleading information. Once the information in question has been determined as misleading information by independent authoritative sources, Microsoft Start's News and Feeds moderation experts create a new policy and guidelines. The new policy then goes through an internal review process before being put into production.

In the six months from October 2021 to March 2022, of the more than 190,000 takedowns in Australia, 2% were due to misinformation traits. A breakdown per trait is in the table below.

Microsoft Start Comments, Australia Takedowns October 2021 - March 2022

Trait	Count	Percentage
Misinformation – COVID	3,353	1.75%
Misinformation – QAnon	265	0.14%
Misinformation – Russia/Ukraine	21	0.01%

Comments data prior to October 2021 is not a reliable metric as the function was only in its early stages. Future reports will provide full reporting period and historical data as available and relevant.

Microsoft Advertising

Microsoft Advertising is our proprietary advertising platform, which serves the vast majority of ads displayed on Microsoft Bing and provides advertising to most other Microsoft services that display ads, as well as many third-party services.

Microsoft Advertising's <u>Misleading Content Policies</u> prohibit advertising content that is misleading, deceptive, fraudulent, or that can be harmful to its users, including



advertisements that contain unsubstantiated claims, or that falsely claim or imply endorsements or affiliations with third party products, services, governmental entities, or organisations.

Microsoft Advertising also has a set of <u>Relevance and Quality Policies</u> to manage the relevancy and quality of the advertisements that it serves through its advertising network. These policies deter advertisers from luring users onto sites using questionable or misleading tactics (e.g., by prohibiting advertisements that lead users to sites that misrepresent the origin or intent of their content).

Under our <u>Sensitive Advertising Policies</u>, Microsoft reserves the right to remove or limit advertising permanently or for a period of time in response to a sensitive tragedy, disaster, death or high profile news event, particularly if the advertising may appear to exploit events for commercial gain or may effect user safety.

In response to the COVID-19 pandemic, in certain instances, related advertising will only be allowed from trusted sources. We prohibit any advertising that exploits the COVID-19 crisis for commercial gain, spreads misinformation or might pose a danger to users' safety. For example, ads and company pages that improperly sell medical supplies and solutions are prohibited.

In response to Russia's invasion of Ukraine in 2022, Microsoft Advertising is preventing serving advertising related to the crisis pursuant to these policies. To enforce our policies, we identified search terms with higher relevancy and user traffic and a list of web domains that violate our policies. We banned all advertisements from Russia Today (RT) and Sputnik across our ad network and will not place any ads from our ad network on these sites. We continue monitoring the situation to update our detection and block mechanisms.

As reported in our <u>2021 in review</u>: Ad safety in <u>Microsoft Advertising</u> blog post, actions taken in 2021 to ensure a safe and trusted experience for users globally included:

- Taking down more than 3 billion ads for various policy violations, almost twice as many as in 2020. We suspended nearly 270,000 accounts and banned nearly 400,000 websites from our network
- Making use of significant advancements in artificial intelligence (AI) to quickly adapt to new patterns and methods used by bad actors
- Ensuring that our protection mechanism involved coverage for all types of content such as text, images, and videos to quickly detect malicious activity in our system
- Making advancements in our human moderation workflows to capture more insights from reviews, continuously improving our systems
- Leveraging intelligent tools to allow our human reviewers to establish linkages between various accounts and discover fraud rings quickly and efficiently

Actions taken during 2020 are reported in our Ad quality year in review 2020 blog post.



Microsoft Advertising Global Ad Takedowns

2020	2021
1.6 billion	3 billion

Microsoft Advertising Ad Safety in Australia, 2021

Action	Global	Australia	
Rejections	3b	191m	
Total appeals	72,413	7,025	
Total appeals overturned	28,965	3,248	
Total complaints	70,000	201	
Main type complaint: Policy violation	20,934	68	
Main type complaint: Trademark infringement	34,700	127	
Main type complaint: User safety issues	416	5	
Other type of complaints	13,950	69	
Total entity takedowns	250,124	2,956	
Average processing time	~36 hours	~36 hours	

In Australia, we have rejected more than 800 ads related to Ukraine, as at mid-April 2022.

In 2021, we introduced <u>advertiser identity verification</u> in seven markets, including Australia.

- This system ensures customers see ads from trusted sources
- The selected advertisers are required to establish their identity as a business or as an individual by submitting all necessary information and documents
- In Australia, 75 accounts have been opted-in to an advertiser identity verification pilot, of which 28 have succeeded, 5 have expired, and 42 are yet to commence

Microsoft strives to ensure that all advertisements on our services are clearly distinguishable from editorial or other non-sponsored content.

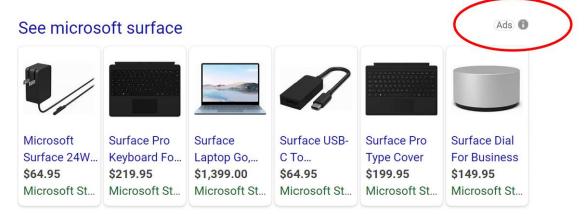
- For example, all Microsoft services that display ads served by Microsoft Advertising clearly distinguish sponsored from non-sponsored content by displaying an advertising label in a readily noticeable location on the page
- An example of how ads are displayed is shown in red below. Clicking on the information icon or downward arrow next to an advertising label displays a click through to the <u>ad setting page</u>



Surface Devices, Accessories - Microsoft Store

https://www.microsoft.com/en-au/store/collections/surfacelist •

The most portable **Surface** touchscreen 2-in-1 is perfect for your everyday tasks, homework and play. Designed to light up the best of Windows 11, **Surface** Go 3 is optimised for digital pen and..



Microsoft Advertising similarly requires all its publishers to use a clear and prominent label indicating that the advertisements served by Microsoft Advertising on their properties are sponsored.

 To enforce this requirement, Microsoft Advertising proactively reviews how its publisher partners integrate Microsoft-served ads into their properties, including an assessment of their compliance with our advertising labelling requirements

LinkedIn

LinkedIn is a real identity professional network. On LinkedIn, the world's professionals come together to find jobs, stay informed, learn new skills, and build productive relationships. The content that our members share becomes part of their professional identity and can be seen by their boss, colleagues, and potential business partners. Accordingly, the content on LinkedIn is professional in nature.

To help keep LinkedIn safe, trusted, and professional, our <u>Professional Community</u> <u>Policies</u> clearly detail the range of objectionable and harmful content that is not allowed on LinkedIn. Fake accounts, misinformation, and inauthentic content are not allowed, and we take active steps to remove it from our platform.

LinkedIn has automated defences to identify and prevent abuse, including inauthentic behaviour, such as spam, phishing and scams, duplicate accounts, fake accounts, and misinformation. Our Trust and Safety teams work every day to identify and restrict inauthentic activity. We're regularly rolling out scalable <u>technologies</u> like machine learning models to keep our platform safe.

As inauthentic behaviour gets more sophisticated, we're improving our detection. Here are some of the latest actions we've taken on fake profiles to help keep our members safe while engaging in our community:



- We have a dedicated team of data scientists, software engineers, machine learning engineers, and investigators who are constantly analysing abusive behaviour on the platform and improving the technology we use to combat it
 - Our team develops automated defences that analyse risk signals and patterns of abuse and take automated action, and constantly improve them to adapt to new threat patterns
- To evolve to the changing threat landscape, our team is investing in new technologies for combating inauthentic behaviour on the platform
 - We are investing in AI technologies such as advanced network algorithms that detect communities of fake accounts through similarities in their content and behaviour, computer vision and natural language processing algorithms for detecting AI-generated elements in fake profiles, anomaly detection of risky behaviours, and deep learning models for detecting sequences of activity that are associated with abusive automation

Using the process described in response to Outcome1c below, LinkedIn members also can report content they believe violates our Professional Community Policies, including misinformation, inauthentic content, and fake accounts. If reported or flagged content violates the Professional Community Policies, it will be removed from the platform. We may also restrict the offending member's LinkedIn account, depending on the severity of the violation and any history of abuse.

In response to the COVID-19 pandemic, <u>LinkedIn editors created and promoted trusted content</u>. LinkedIn introduced the following measures:

- Promote content from most credible organisations and experts, such as the World Health Organization
- Launched a 'Special Report: Coronavirus' box above 'Today's News and Views' with story lines relevant to COVID-19 and including updates from the World Health Organization (WHO) and Centers for Disease Control and Prevention (CDC)

LinkedIn's team of global editors also proactively provide members with curated news from trusted sources in a number of languages, and its content moderation teams closely monitor associated conversations in languages including Russian, Ukrainian, German, Dutch, and French.

- For example, LinkedIn's team of editors cover the most recent developments of Russia's invasion of Ukraine, ranging from the economic impact to major military events that are taking place
- These news reports are prominently displayed in the member's content feed and in banner notifications. In addition, LinkedIn directs members searching for information about the conflict to news from these trusted sources

LinkedIn had a dedicated, direct means of communication with the Australian Electoral Commission for content referrals in breach of the Electoral Act during the 2022 Australian Federal Election campaign period.



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.

Users can find information about the types of behaviours and content that will be prohibited and/or managed as follows:

- Microsoft Advertising: Microsoft Advertising policies
- Microsoft Start: Microsoft Services Agreement, Community Guidelines
- LinkedIn: Professional Community Policies

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

In addition to the guidelines contained within the respective user agreements, **Microsoft Bing, Microsoft Start, Microsoft Advertising** and **LinkedIn** have reporting mechanisms where users are able to flag problematic content.

Microsoft Bing

Microsoft Bing users can use the feedback tool (see below) to select a specific search result to flag. The user can include a screenshot of the specific search result, as well as a description of the issue.

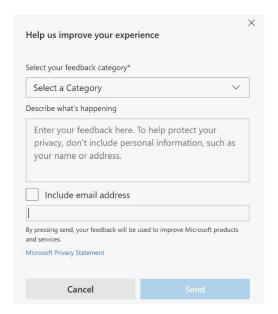
Microsoft Bing users can also use the <u>Report a Concern</u> page to directly report other types of content, as well as concerns about misinformation and disinformation through the 'other concern' option.

Feedback about Bing Please click on the specific area of the page that	Tell us about your concern.
your feedback is related to so it can be sent to the correct team.	A broken link or outdated page
Suggest	O Intellectual property (copyright, trademark, sale of counterfeit goods)
Like	Child sexual exploitation and abuse imagery ("child pornography")
Dislike	Offensive material
	My private information (intimate or sexual imagery, credit card numbers, passwords)
Enter your feedback here. To help protect your	Minor child image(s) of me or someone I am legally authorized to represent
privacy, don't include personal info, like your name or email address.	○ I have a court order
fullic of circuit address.	Malicious pages (phishing, malware)
✓ Include a screenshot	I have another concern

Microsoft Start

Microsoft Start includes a feedback feature at the bottom of all pages (landing page and each article, see below), with Content Quality as one of the options in the drop-down menu. This feedback feature is also included in the Settings menu.





Microsoft Advertising

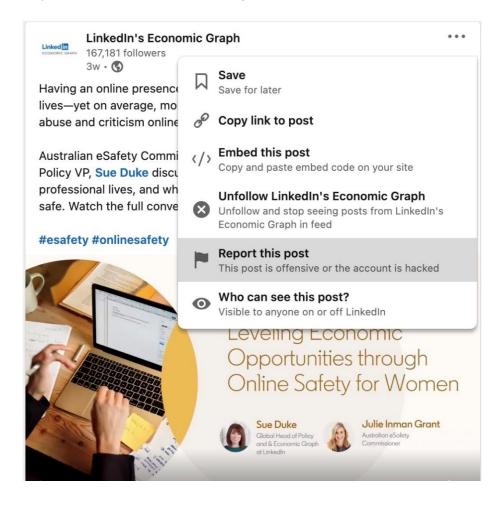
Microsoft Advertising enables users to report ads which may be in violation of its policies (e.g., ads that may contain malvertising, disallowed content, relevancy concerns, or sensitive content) through its <u>low quality ad submission and escalation form</u>. Users of Microsoft Bing and Microsoft Start can also report ads via the respective feedback functions on those services (see below).



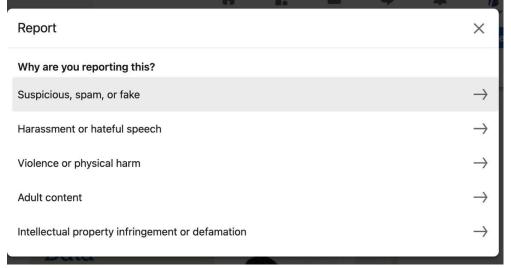


LinkedIn

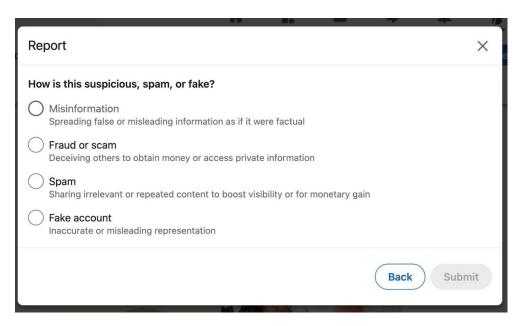
If **LinkedIn's** members locate content they believe violates our <u>Professional Community Policies</u>, we encourage them to report it using the in-product reporting mechanism represented by the three dots in the upper right hand corner of a post on LinkedIn:

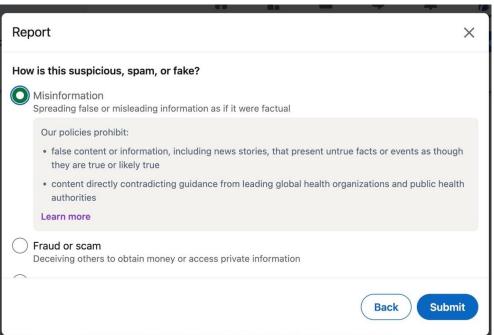


Misinformation is specifically called out as one of the reporting options:

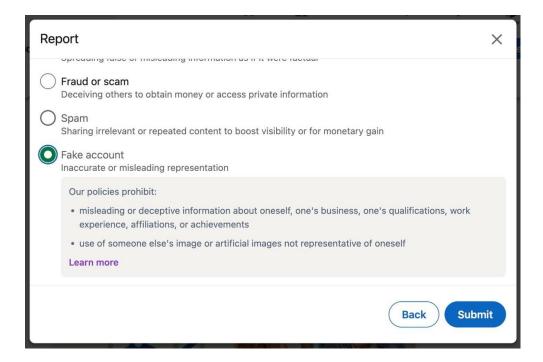












As shown above, LinkedIn has a robust reporting mechanism that allows users to report content they believe violates our policies, including misinformation and fake accounts.

- We encourage our members to report abusive content, messages, or safety concerns, whether in profiles, posts, messages, comments, or anywhere else
- Once flagged, our system sends the content to our team of content reviewers for manual review
- Our review team evaluates this content to determine if, in fact, it violates our policies and, if so, removes the content where appropriate
- We may also restrict the offending member's account, depending on the severity of the violation and any history of abuse

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

Microsoft Bing, Microsoft Start, Microsoft Advertising

Microsoft regularly publishes information about the detection and removal of content that violates our policies or is subject to removal under local legal obligations in the Digital Trust section of our Reports Hub. This includes

- <u>Content Removal Requests Report</u>, which includes government requests for content removal (Australia did not feature in reporting period), copyright requests (global total) and right to be forgotten (EU only)
- <u>Digital Safety Content Report</u>, which provides information on actions taken in relation to child sexual exploitation and assault imagery (CSEAI), terrorist and violent extremist content (TVEC) and non-consensual intimate imagery (NCII)



Moving forward, Microsoft plans to create regular and consistent data reporting to support further transparency in our services. Where possible, this reporting will provide greater insights across geographic regions and countries.

In addition, <u>2021 in review: Ad safety in Microsoft Advertising</u> blog post includes data on global complaints received and addressed, including type of complaint, takedowns, appeals, average processing time. It also addresses focus areas of the reporting year and going forward.

LinkedIn

The **LinkedIn** Community Report describes actions we take on content that violates our Professional Community Policies and User Agreement. It is published twice per year and covers the global detection of fake accounts, spam and scams, content violations and copyright infringements. In the most recently reporting period from 1 January to 30 June 2021:

- 11.6 million fake accounts were stopped at registration
- 3.7 million fake accounts were restricted proactively before members reported
- 85,700 fake accounts were restricted after members reported
- 97.1% of fake accounts were detected by automated defences and 2.9% manually
- 66.1 million scams/spam were proactively detected and removed by LinkedIn
- 232,000 scams/spam were removed after member reports
- 147,490 pieces of misinformation content were removed by LinkedIn
 - Misinformation was the largest category of content violation, ahead of harassment or abusive in this reporting period

The Community Report for July to December 2021 will be available <u>here</u> in the near future, and data will be included in next year's report.

LinkedIn Community Report: global actions taken on content that violated Professional Community Policies and User Agreement, January 2019 – June 2021

-		Global			Australia		
		2019 Jan – Jun	2019 Jul – Dec	2020 Jan – Jun	2020 Jul – Dec	2021 Jan - Jun	2021 Jan – Jun
Fake Accounts	Stopped at registration	19.5m	7.8m	33.7m	11.6m	11.6m	54,883
	Restricted proactively before members reported	2m	3.4m	3.1m	3.0m	3.7m	64,642
	Restricted after members reported	67.4k	85.6k	103.1k	111k	85.7k	1,281
Content Violations	Misinformation*			22.8k	110.7k	147.5k	2,149



*Misinformation not reported as a separate category prior to 2020. Other content violation categories reported are harassment, adult, hateful or derogatory, violent or graphic, child exploitation.

Objective 2: Disrupt advertising and monetisation incentives for Disinformation.

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

Microsoft strives to provide our customers with a positive online experience free from deceptive advertisements. Microsoft is working across our services to achieve this goal through policies and enforcement processes aimed at ensuring that the advertising and content served is clear, truthful, and accurate.

Microsoft Advertising

Advertisers

Microsoft Advertising's policies on <u>Misleading Content</u>, <u>Relevance and Quality</u>, and <u>Sensitive Advertising</u> disrupt and reduce advertising and or monetisation incentives for disinformation.

Microsoft Advertising employs dedicated operational support and engineering resources to enforce these policies:

- Automated and manual enforcement methods are used to prevent or take down advertisements that violate policy
- Manual review of all advertisements flagged to the customer support team
- Customer support team seeks feedback from advertisers and investigates their complaints

Every ad loaded into the Microsoft Advertising system is subject to these enforcement methods, which leverage machine-learning techniques, automated screening, the expertise of its operations team, and dedicated user safety experts.

Publishers

Regarding publishers, Microsoft Advertising uses a distinct set of measures designed to avoid the display of advertising on—and thus disrupt the flow of advertising revenue to—sites involved in spreading disinformation.

Microsoft Advertising works with selected, trustworthy publishing partners who are required to:

- Not serve ads that contain sensitive political content (e.g., extreme, aggressive, or misleading interpretations of news, events, or individuals), unmoderated usergenerated content, and unsavoury content (e.g., disparaging individuals or organisations)
- Maintain a list of prohibited terms and provide information on content management practices



• Abide by restrictions against engaging in business practices that are harmful to users (e.g., distributing malware)

In monitoring publishers to ensure reduction and disruption of disinformation and misinformation, Microsoft Advertising:

- Reviews publisher properties and domains for policy compliance, including compliance with restrictions on prohibited content
- Notifies publishers promptly if properties or domains violate policies and does not approve for live ad traffic
- Removes live property or domain found in violation until the publisher remedies the issue

Publishers also benefit from the set of measures identified above that Microsoft Advertising takes with regard to advertisers, which ensures that these partners receive high-integrity, non-deceptive ads from the Microsoft Advertising platform.

Microsoft Advertising's action in response to Russia's invasion of Ukraine is outlined in response to Outcome 1a.

Changes to processes and circumstances globally are detailed in the <u>2021 in review</u>: Ad <u>safety in Microsoft Advertising</u> blog post.

LinkedIn

LinkedIn has a number of policies and processes in place to disrupt attempts to publish advertising with disinformation.

All content on LinkedIn must comply with the <u>Professional Community Policies</u>, which prohibit, among other things, disinformation and misinformation.

In addition, advertising on LinkedIn must also comply with LinkedIn's <u>Advertising Policies</u>, which prohibit the following:

- Fraudulent and deceptive ads. Any claims in an ad must have factual support
- Political ads. This includes ads advocating for or against a particular candidate, party, or ballot proposition or otherwise intended to influence an election outcome; ads fundraising for or by political candidates, parties, political action committees or similar organisations, or ballot propositions; and ads exploiting a sensitive political issue even if the advertiser has no explicit political agenda
- Ads that attempt to exploit or that are inappropriate during sensitive events, tragedies, and disasters
- Ads that make unrealistic or misleading claims about health
- Ads concerning illegal products, services, and activities

To help ensure LinkedIn remains a safe, trusted, and professional platform, LinkedIn reviews submitted ads against its policies using a combination of automated defences and human review and applies its policies to keep ads containing disinformation and misinformation off LinkedIn, including, for example, regarding COVID-19 and Russia's invasion of Ukraine.



LinkedIn also makes it easy for members to <u>report</u> ads they believe may violate LinkedIn policies.

Objective 3: Work to ensure the security and integrity 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.

In addition to the actions detailed in Objective 1 (Outcomes 1a, 1b and 1c), **Microsoft Bing, Microsoft Advertising**, and **LinkedIn** to reduce the risk of inauthentic user behaviours through the measures detailed below.

Microsoft Bing

Users cannot post content to **Microsoft Bing** directly which means it is less vulnerable to certain types of harmful bot and false account activity that afflict many other types of online services (e.g. social media services).

Microsoft Bing search results draw on an index of third-party web content created by Microsoft itself by crawling the open web, which are ranked using proprietary algorithms. Users cannot share or "like" content or otherwise cause content to go "viral". This limits substantially the possibility for many types of abuse common to social media platforms.

That does not mean that bots do not pose a threat to the service, and Microsoft Bing takes significant efforts to ward off other types of bot activity.

- For example, Microsoft Bing observes attempts to abuse its search platform by bots seeking to influence its ranking system (e.g., by engaging in "click fraud" or harnessing "link farms" to make it seem as though a site is more popular than it actually is)
- Microsoft Bing's <u>Webmaster Guidelines</u> provide an overview of the types of inauthentic behaviours that are likely to lead to ranking penalties

Microsoft Advertising

Microsoft Advertising employs a robust filtration system to detect bot traffic.

- This system uses various algorithms to automatically detect and neutralise invalid or malicious online traffic which may arise from or result in click fraud, phishing, malware, or account compromise
- The system is supported by several teams of security engineers, support agents, and traffic quality professionals who continually develop and improve monitoring and filtration
- Support teams work closely with advertisers to review complaints around suspicious online activity and across internal teams to verify data accuracy and integrity

LinkedIn

As a real identity professional network, **LinkedIn** acts vigilantly to maintain the integrity of all accounts and to ward off bot and false account activity. LinkedIn enforces the policies in its



<u>User Agreement</u> prohibiting the use of "bots or other automated methods to access the Services, add or download contacts, send or redirect messages" through:

- Having a dedicated Anti-Abuse team to create the tools to enforce this prohibition
- Using automated systems detect and block automated activity
- Imposing hard limits on certain categories of activity commonly engaged in by bad actors
- Detecting whether members have installed known prohibited automation software
- Conducting manual investigation and restriction of accounts engaged in automated activity
- Partnering with the broader Microsoft organisation to develop technological solutions for detecting "deep fakes"
- Investing in and using AI to detect coordinated inauthentic activity and communities
 of fake accounts through similarities in their content and behaviour

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.

Microsoft is committed to helping our customers make informed decisions about content. This includes providing our customers with tools to help them evaluate the trustworthiness of that content.

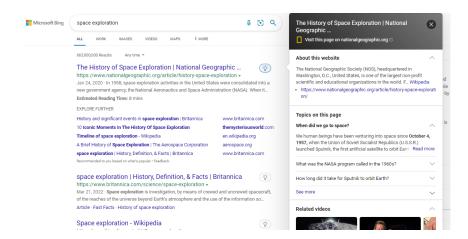
Microsoft is working both internally and with third parties to provide new tools and implement new technologies across our services to assist our customers in identifying trustworthy, relevant, authentic, and diverse content, including in news, search results, and user-generated material.

Microsoft Bing

In addition to the way **Microsoft Bing** ranks content to promote relevant, quality, and credible material, Microsoft Bing has created tools to help users independently evaluate the legitimacy of content they find online through Microsoft Bing search results. Microsoft Bing is continually refining its algorithms to ensure it returns the highest authority results possible to its users.

Microsoft Bing has developed a Page Insights feature that provides additional information about the website providing the search result through links to Wikipedia or similar sources. An example of the Page Insights feature is below.





Microsoft Bing offers an "Intelligent Search" feature in response to certain user queries that demonstrate an intent by a user to learn all valid answers to a question.

- Identifying questions with multiple valid answers involves several techniques, including sentiment analysis, which identifies the opinions expressed in a piece of text (i.e., positive, negative, or neutral)
- Through this feature, Microsoft Bing will summarise the various potentially valid answers to a user query and display them all in a carousel to give the user a balanced overview

Microsoft Bing also empowers consumers by continuously improving its ability to accurately and instantaneously rank sites along a high/low authority continuum.

- Research indicates that websites that promulgate disinformation tend to meet Microsoft Bing's internal classification standards for "low authority" sites
- As Microsoft Bing gets better and better at differentiating high from low authority material, consumers will be exposed to fewer sites of low reliability (including sites disseminating disinformation) when searching for information and content online

Microsoft Start

In response to the invasion of Ukraine, **Microsoft Start** has added a "War in Ukraine" heading topic, to enable users to easily access trustworthy content on the issue.



Microsoft Start clearly labels advertising to enable users to readily distinguish this from other content.

LinkedIn

LinkedIn has implemented multiple programs to empower users to make better informed choices about digital content.

• LinkedIn has an internal team of experienced news editors that provides trustworthy and credible news about current events to the LinkedIn member base. LinkedIn also creates numerous entry points to this information in the form of feed promotions,



- notifications, and a banner that points to this information when members search for related topics
- Curated, interest-based feeds provide members with news and conversations most relevant to them and their industry
- LinkedIn also shows members how to delete conversations or turn off comments if
 they feel unsafe or feel that conversation may be steered towards being unproductive
 or unprofessional. Members can also choose their audience when they post on
 LinkedIn, selecting who gets to see your content
- To help guide and members, LinkedIn also maintains "Best practices" guidelines for sharing quality content
- The <u>Influencer Program</u> handpicks over 500 leaders in their respective industries to create and share reliable, trustworthy content

Other contributions and measures

Globally, Microsoft also has a number of programs to proactively combat disinformation on our services and empower users.

Microsoft expects that methods for generating synthetic media, or "deep fakes", which are photos, videos or audio files manipulated by AI in hard-to-detect ways, will continue to grow in sophistication.

- As all Al detection methods have rates of failure, platforms must understand and be ready to respond to deep fakes that slip through detection methods
- In the longer term, there will be a need for stronger methods for maintaining and certifying the authenticity of news articles and other media
- There are few tools today to help assure readers that the media they are seeing online came from a trusted source and that it was not altered

Microsoft is devoting innovation and research resources to this problem.

Video Authenticator

Video Authenticator was created to address deep fakes.

- Video Authenticator can analyse a still photo or video to provide a percentage chance, or confidence score, that the media is artificially manipulated
- In the case of a video, it can provide this percentage in real-time on each frame as the video plays
- It works by detecting the blending boundary of the deep fake and subtle fading or greyscale elements that might not be detectable by the human eye

Project Origin and Coalition for Content Provenance and Authenticity (C2PA)

Microsoft has developed <u>Project Origin</u>, which aims to create a measure of accountability for media to reduce the spread of disinformation through synthetic media. This technology helps certify the source of the content, like a watermark, which will expose if alterations have been made which can lead to manipulation and disinformation.

 Project Origin gives publishers and consumers a tool to identify when media has been altered from the original source



- While it is ultimately up to viewers to believe in the accuracy or truth of a publisher, knowing if media has been tampered with gives us more power to determine what we can trust
- As part of the publication process, tools will register media items by creating a digital fingerprint
 - In return, content creator will receive a certification of authentication, which will be stored in a tamper-proof distributed ledger with no single controlling entity
 - o The certification can be embedded into a piece of media before distribution
- When a user consumes a piece of media, the web browser or dedicated application
 will automatically compare the embedded digital fingerprint of the file being viewed
 with the original stored in the distributed ledger
 - Based on that comparison, it will display a clear indicator of authentication, allowing users to understand if what they are viewing is as published or has been tampered with

Microsoft is a founding member of the <u>Coalition for Content Provenance and Authenticity</u> (C2PA).

- The Coalition aims to address the prevalence of disinformation, misinformation, and online content fraud through developing technical standards for certifying the source and history or provenance of media content
- Standards are still in development, as member organisations work together to develop content provenance specifications for common asset types and formats to enable publishers, creators, and consumers to trace the origin and evolution of a piece of media, including images, videos, audio, and documents
 - These technical specifications will include defining what information is associated with each type of asset, how that information is presented and stored, and how evidence of tampering can be identified

Please also see response to Objective 6.

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.

Microsoft Advertising

Under our <u>Advertising Policies</u>, **Microsoft Advertising** prohibits political advertising. This includes ads for election-related content, political candidates, parties, ballot measures, and political fundraising globally; similarly, ads aimed at fundraising for political candidates, parties, political action committees ("PACs"), and ballot measures also are barred.

All Microsoft and third-party services that rely on Microsoft Advertising to serve advertisements on their platforms benefit from these robust, and robustly enforced, set of policies.



Specifically, Microsoft Advertising employs dedicated operational support and engineering resources to enforce restrictions on political advertising using a combination of proactive and reactive mechanisms.

- On the proactive side, Microsoft Advertising has implemented several processes designed to block political ads from showing across its advertising network, including restrictions on certain terms and from certain domains
- On the reactive side, if Microsoft Advertising becomes aware that an ad suspected of violating its policies is being served to our publishers—for instance, because someone has flagged that ad to our customer support team—the offending ad is promptly reviewed and, if it violates our policies, taken down

Microsoft Advertising's policies also prohibit certain types of advertisements that might be considered issue based. More specifically, "advertising that exploits political agendas, sensitive political issues or uses 'hot button' political issues or names of prominent politicians is not allowed regardless of whether the advertiser has a political agenda," and "advertising that exploits sensitive political or religious issues for commercial gain or promote extreme political or extreme religious agendas or any known associations with hate, criminal or terrorist activities" is also prohibited.

LinkedIn

LinkedIn does not accept political advertising. LinkedIn's <u>Advertising Policies</u> globally prohibit political ads which:

- advocate for or against a particular candidate, party or ballot proposition or are otherwise intended to influence an election outcome
- fundraise for or by political candidates, parties, ballot propositions or political action committees or similar organisations
- exploit a sensitive political issue even if the advertiser has no explicit political agenda

All ads are subject to review for adherence to policy before being approved to run. LinkedIn has also introduced features making it simple for members to report advertisements that violate LinkedIn's policies; LinkedIn reviews such reports and removes offending advertisements from its platform.

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.

Microsoft is working at a company-wide level to ensure that the research community has the tools and resources it needs to play its crucial part in helping to combat disinformation. A non-exhaustive list of Microsoft's ongoing collaborations with the broader research community in this space include:

Microsoft Research on	Teams within Microsoft Research are collaborating closely
COVID-19	with Professor Jacob Shapiro, director of the Empirical



	Studies of Conflict Project at Princeton University, to characterise the types and extent of misinformation and disinformation narratives online related to COVID-19.
	This helps our researchers, product teams, and industry partners understand the global information environment our customers are exposed to.
Technology Academics Policy (TAP)	TAP is a forum for leading academics to share articles, ideas and research that focus on the impact of technological change with each other and the broader online community.
	Microsoft provides administrative and financial support.
Oxford Technology and	Microsoft supports OxTEC to research how democracies can
Elections Commission	integrate democratic norms and practices into the use of
(OxTEC), Oxford Internet Institute	information technologies, social media, and big data during campaigns, with the goal of protecting the integrity of elections.
Trust Project	Microsoft Start, then Bing News, joined the Trust Project in 2017. The consortium led through Santa Clara University's Markkula Center for Applied Ethics, to develop a standardised technical language for platforms to learn more from new sites about quality and expertise behind journalists' work.
	This Project includes organisations such as the Washington Post, the Economist, and the CBC. It does not yet list any Australian partners.
	The Trust Project is the first to give search engines and social media platforms the consistent technical standards they need to surface reliable, relevant, and honest news. Microsoft Bing uses the Trust Indicators in display and behind the scenes in some markets.
Microsoft Research Lab	Microsoft brings together social scientists and humanists from the fields of anthropology, communication, economics, information, law, media studies, women's studies, science & technology studies, and sociology to provide insight into how social media is reconfiguring society.
	Microsoft Research also maintains the Social Media Collective, a network of social science and humanistic researchers with purpose to enhance understanding of the social and cultural dynamics that underpin social media technologies.



Partnership on Al	Microsoft is a partner in Partnership on AI which works to
	better understand and address the emerging threat posed
	by the use of AI tools to develop malicious synthetic media
	(i.e., deep fakes).
Partnership with Miburo	Microsoft partner Miburo has published research on
	Russian disinformation, including information on targets
	and tactics being used since the invasion of Ukraine and
	analysis of Russia's propaganda and disinformation
	<u>ecosystem</u> .
MS MARCO	Microsoft Bing makes information available to the research
	community to improve search results by making data sets
	like MS MARCO publicly available.

Democracy Forward Program

Microsoft's Democracy Forward Program (then known as the Defending Democracy Program) was launched in 2018. It is an innovative effort to protect democratic institutions and processes from hacking, to explore technological solutions to protect electoral processes, and to defend against disinformation.

Microsoft believes technology companies have a responsibility to help protect democratic processes and institutions globally. The Democracy Forward Program also leverages Microsoft's role as a business and software provider to increase our clients' ability to counter outside efforts to compromise their security infrastructure. While these services do not target disinformation directly, they can help reduce its spread by increasing security against bad actors. These services include:

- AccountGuard, a free service for eligible Outlook 365 customers involved in elections, such as political parties and campaigns, political vendors, and think tanks
 - AccountGuard provides notifications about cyber threats, guidance on best practices for dealing with the unique problems faced by politically oriented organisations, and access to security features typically offered only to large corporate and government account customers
 - AccountGuard is available in 32 countries, including in Australia. It is
 protecting more than 4.6 million inboxes from political campaigns, election
 officials, political parties, political consultants, think tanks, democracy
 advocacy organisations, human rights organisations, journalists and
 newsrooms, the healthcare industry, and non-profit organisations
- Cybersecurity trainings offered first-party by the Democracy Forward team, which have been attended by over 2,000 individuals from political parties, political campaigns, election administration and supply chain, non-profit organisations, and think tanks
- Security trainings to which we have contributed have had 8,000 individual participants globally
- <u>ElectionGuard</u> and <u>Microsoft 365 for Campaigns</u> promote election integrity and campaign security



Partnership with NewsGuard

Microsoft also helps its customers evaluate the quality of the news they encounter on the Internet through its partnership with <u>NewsGuard Technologies</u>.

- Led by respected journalists and entrepreneurs Steven Brill and Gordon Crovitz, NewsGuard's analysts review online news sites across a series of nine journalistic integrity criteria
 - Criteria include whether the site regularly publishes false content, reveals conflicts of interest, discloses financing, and publicly corrects reporting errors.
- NewsGuard then compiles their findings into a "Nutrition Label" and corresponding Red/Green Reliability Rating, which users can view to better understand the reliability of the news they consume
- NewsGuard is currently available in the US, UK, Canada, France, Germany, and Italy

Microsoft has partnered with NewsGuard to provide a free plug-in for the Microsoft Edge web browser (also available for other browsers including Chrome and Firefox), as well as an opt-in news rating feature for the Edge mobile application on both iOS and Android. This empowers Edge users to benefit from the comprehensive analysis done by NewsGuard and to better identify the most reliable news and information sites.

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

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

Our reporting under this code is available on the Microsoft Australia News Centre.

Microsoft also releases other information about our initiatives globally to combat disinformation:

Microsoft On the Issues	Blog contains announcements on technology policy	
	issues, including disinformation.	
	For example, our response to the invasion of Ukraine,	
	COVID-19, video authenticator technology, release of	
	digital trust reports, are all posted on the blog	
Microsoft Reports Hub	Digital Trust reports (Content Removal, Digital Safety,	
	Law Enforcement Requests)	
	Privacy Report	
	Standards of Business Conduct	
Microsoft Advertising Blog	Ad quality annual year in review post	
Microsoft Digital Defense	Report encompasses learnings from security experts,	
Report	practitioners, and defenders at Microsoft to empower	
	people everywhere to defend against cyberthreats.	
	Includes dedicated section on disinformation.	



LinkedIn Transparency	Community Report
Report	Government Requests Report
<u>LinkedIn Blog</u>	Blog contains information on actions to combat
	disinformation, including How We're Protecting
	Members From Fake Profiles, Automated Fake Account
	Detection, and An Update on How We Keep Members
	Safe

Conclusion

Microsoft is committed to playing our part in stamping out disinformation through application of our policies, our own research and innovation in new technologies, and collaboration with partners, academia, and our users. This report details some of the initiatives Microsoft Bing, Microsoft Start, Microsoft Advertising and LinkedIn are taking to reduce and disrupt the propagation of disinformation and misinformation, as well as the efforts the company is taking to contribute to the goals and commitments of the Australian Voluntary Code of Practice on Disinformation and Misinformation.