Microsoft

Australian Code of Practice on Disinformation and Misinformation

Microsoft and LinkedIn

Annual Transparency Report

May 2026

Summary

Microsoft is pleased to file this report on our commitments under the voluntary Australian Code of Practice on Disinformation and Misinformation (the **Code**), covering the reporting period of calendar year 2025.

We have submitted Transparency Reports under the Code every year since 2021. In each Transparency Report, we have shown how Microsoft is committed to instilling trust and security across our products and services, and across the broader online ecosystem. We continue to recognise that fighting disinformation is a key element to creating a trustworthy and safe online environment and continue to increase our efforts to counter these threats.

We also recognise that there is not a one size fits all approach to this work, and instead there needs to be a whole of society strategy that recognises that not all people or platforms are the same and that different measures may be more effective than others in improving the information environment.

The Microsoft services in scope of the Code are:

- **Microsoft Advertising:** Microsoft's proprietary online advertising network, which serves ads displayed on Bing Search and provides advertising to most other Microsoft services that display ads.
- **Bing Search:** a web search engine which provides a variety of services including web, video, image, and map search products. Bing Search 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.
- **MSN (formerly known as Microsoft Start):** a service which delivers licenced news and content across web and mobile on behalf of Microsoft customers and syndication partners.
- **LinkedIn:** a real-identity online networking service for professionals to establish their professional identities online, connect and interact with other professionals, and build relationships for the purpose of collaborating, learning, and staying informed about industry information and trends. It operates via websites and mobile apps and includes user-generated content.

Commitments under the Code

Commitment	Relevant Microsoft service
1a: Contribute to reducing risk of harm by adopting scalable measures	Bing Search, MSN, Microsoft Advertising, LinkedIn
1b: Users informed about types of behaviours and content prohibited/managed	MSN, Microsoft Advertising, LinkedIn
1c: Users can report content that violates policy through accessible reporting tools	Bing Search, MSN, Microsoft Advertising, LinkedIn
1d: Users can access general information about response	Bing Search, MSN, LinkedIn
1e: Users will be able to access general information about use of recommender systems and have options related to content suggested by recommender systems	LinkedIn
2: Advertising and/or monetisation incentives reduced	Microsoft Advertising, LinkedIn
3: Risk of inauthentic behaviours undermining integrity and security of services/products reduced	Bing Search, Microsoft Advertising, LinkedIn
4: Users are enabled to make informed choices about sources of news and factual content and to identify misinformation	Bing Search, MSN, 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	Bing Search, MSN, Microsoft Advertising, LinkedIn

Unless otherwise specified, data provided in this Transparency Report is for 2025 calendar year.

Microsoft Advertising

Microsoft Advertising works both with advertisers, who provide it with advertising content, and publishers, such as Bing Search, who display these advertisements on their services. Microsoft Advertising employs a distinct set of policies and enforcement measures with respect to each of these two categories of business partners to prevent the spread of disinformation through advertising.

Bing Search

Bing Search is an online search engine with the primary objective of connecting users with the most relevant search results from the web. Users come to Bing with a specific research topic in mind and expect Bing to provide links to the most relevant and authoritative third-party websites on the internet that are responsive to their search terms. Therefore, addressing misinformation in organic search results often requires a different approach than may be appropriate for other types of online services. Blocking content in organic search results based solely on the truth or falsity of the content can raise significant concerns relating to fundamental rights of freedom of expression and the freedom to receive and impart information.

While Bing's efforts may on occasion involve removal of content from search results (where legal or policy considerations warrant removal), in many cases, Bing has found that actions such as targeted ranking interventions, or additional digital literacy features such as Answers pointing to high authority sources, trustworthiness signals, or content provenance indicators, are more effective. Bing regularly reviews the efficacy of its measures to identify additional areas for improvement and works with internal and external subject matter experts in key policy areas to identify new threat vectors or improved mechanisms to help prevent users from being unexpectedly exposed to harmful content in search results that they did not expressly seek to find.

Bing offers generative AI experiences, including Copilot Search, Bing Image Creator, and Bing Video Creator. Copilot Search utilises AI to deliver a unique experience by not only optimising search results but presenting information in a user-friendly, cohesive layout. Results also include citations and links that enable users to explore further and evaluate websites for themselves. For these AI-powered experiences, Bing partnered closely with Microsoft's Responsible AI team to proactively address AI-related risks and continues to evolve these features based on user and external stakeholder feedback.

LinkedIn

LinkedIn's vision is to create economic opportunity for every member of the global workforce. Its mission is to connect the world's professionals to make them more productive and successful. LinkedIn is a networking tool that enables members to establish their professional identities online, connect with other professionals, and build relationships for

the purpose of collaborating, learning, and staying informed about industry information and trends.

LinkedIn is a real-identity platform, where members must use their real or preferred *professional* names, and the content they post is visible, for example, to their colleagues, employers, potential future employers, and business partners. Given this audience, members by and large tend to limit their activity to professional areas of interest and expect the content they see to be professional in nature.

LinkedIn is focused on keeping its platform safe, trusted, and professional, and respects the laws that apply to its services. On joining LinkedIn, members agree to abide by LinkedIn's User Agreement and its Professional Community Policies, which expressly prohibit the posting of information that is intentionally deceptive or misleading.

LinkedIn uses systems, technology, and reports from its members to detect and quickly remove content that violates LinkedIn's Professional Community Policies. In 2025, LinkedIn globally blocked more than 197 million fake accounts (a majority of which were stopped at registration) and removed 44,564 pieces of misinformation. Over the same period, LinkedIn blocked more than 1,400,000 fake accounts attributed to Australia and removed 933 pieces of misinformation reported, posted, or shared by Australian members.

MSN

MSN is a personalised feed of news and informational content from publishers available in a number of Microsoft products, including a standalone website (MSN.com), a mobile app on both Android and iOS, the News and Interests experience on the Windows 10 taskbar, the Widgets experience in Windows 11, and the Microsoft Edge new tab page. On MSN, we have policies to specifically address disinformation and misinformation on clear and well-defined misinformation narratives.

Our approach

Microsoft announced the first Information Integrity Principles in 2022. These principles continue to be adopted across all impacted Microsoft products and teams to ensure an enterprise approach to information integrity while also recognising the immense diversity across the company. The four information integrity principles are:

- **Freedom of Expression**: We will respect freedom of expression and uphold our customers' ability to create, publish, and search for information via our platforms, products, and services.
- **Authoritative Content:** We will prioritise surfacing content to counter foreign cyber influence operations by utilising internal and trusted third-party data on our products.
- **Demonetisation**: We will not wilfully profit from foreign cyber influence content or actors.
- **Proactive Efforts:** We will proactively work to prevent our platforms and products from being used to amplify foreign cyber influence sites and content.

Combating harmful and deceptive AI

The focus on artificial intelligence (**AI**) and interest in understanding how AI could affect the spread of disinformation continues to grow. While AI certainly poses challenges in the information integrity space, we continue to see many opportunities for AI to assist and streamline defenders' work in detecting and assessing influence operations. To be clear, challenges include the evolving tactics and potential efforts to create or disseminate malicious content. However, Microsoft is fully committed to utilising best in class tools, practices, and technology to help mitigate the risks of its services being used to further disinformation. Serving as a leader in AI research, we are committed to proactively publicize our threat detection efforts for the benefit of society. As such, we have adopted six focus areas to combat the harmful use of deceptive AI:

1. A strong safety architecture
2. Durable media provenance and watermarking
3. Safeguarding our services from abusive content and conduct
4. Robust collaboration across industry and with governments and civil society
5. Modernised legislation to protect people from the abuse of technology
6. Public awareness and education

Microsoft continues to invest in partnerships and products that advance public awareness, education, and resilience against deceptive AI globally. Through the global human rights organisation WITNESS, Microsoft supports frontline journalists, fact-checkers, and civil

society actors in detecting and assessing the authenticity of content and the deceptive use of generative AI.

In education, Microsoft's Reed Smart experiential learning game on Minecraft Education teaches students to investigate cases of AI misuse through gameplay, reaching millions of learners globally, including in Australia. Additionally, Search Progress, an information literacy learning accelerator embedded in Microsoft Education tools, helps students build the critical evaluation skills needed to navigate an information environment increasingly shaped by AI. Updates to Search Progress released in the first quarter of 2026 enhance features designed to help students assess sources, cross-check information, and reason through questions of source credibility. Search Progress has been adopted in more than ten countries.

Election integrity

Since the adoption of the Tech Accord to Combat Deceptive Use of AI in 2024 Elections (Tech Accord), Microsoft has made significant strides to prevent the creation of deceptive AI targeting elections, enhance detection and response capabilities, and improve transparency and public awareness. Since that time, Microsoft has continued to work to make it more difficult for malicious threat actors to use legitimate tools to create deceptive AI-generated content targeting democracies around the world while simultaneously simplifying the process for users to identify authentic content.

We empower candidates, campaigns and election authorities to help us detect and respond to deceptive AI targeting elections. Microsoft maintains a reporting form for candidates, campaigns, and election authorities to directly report deceptive AI content on Microsoft consumer services. This reporting tool allows for 24/7 reporting by impacted election entities who have been targeted by deceptive AI found on Microsoft platforms.

Additionally, Microsoft is committed to advancing trusted information and believes that including content credentials is an important driver for this. We were a founding member of the Coalition for Content Provenance and Authenticity (C2PA). To achieve transparency, support information integrity, and empower our users, we leverage C2PA's "content credentials" open standard across several products. Content containing the "Content Credentials" technology will be automatically labelled on LinkedIn, with users seeing the "Cr" label. By clicking the label, users will be able to trace the origin of the media, including the source and history of the content, and whether it was created or edited by AI.

In addition, Microsoft launched the AI & Elections Accelerator, a hands-on training program to help election offices around the world understand AI tools, potential use cases in election administration, and effective strategies to mitigate risks. Microsoft worked with partner International IDEA to help scale this program globally, reaching over 1,000 election officials from 80 countries in the first nine months of the program alone.

In Australia, Microsoft is an inaugural signatory to the Electoral Council of Australia and New Zealand (ECANZ) Statement of Intent with Online Platforms, designed to support electoral

management bodies and online platforms to work together to promote and support the integrity of electoral events in Australia and New Zealand.

Microsoft's role in the 2025 Australian Election

1. **Operational Continuity and Security**: Microsoft provided heightened technical monitoring and support for resilience of cloud and other services delivering electoral systems and other important workloads throughout the election period. Our extensive security monitoring capabilities were also brought to bear. We provided heightened threat monitoring for Australian agencies and other customers who we know are more likely targeted by threat actors during election periods. Our Microsoft Threat Analysis Center (MTAC), Microsoft Threat Intelligence Center (MSTIC), Digital Crimes Unit (DCU) and Global Hunting Oversight and Strategic Triage (GHOST) teams worked to identify foreign influence operations, cybersecurity threat activities, financial crime and AI-driven disinformation campaigns targeted at Australia.
2. **Public Engagement:** In January 2025, Microsoft's Elections and Societal Resilience experts met with more than 150 people from Australia's political parties and their candidates, newsrooms, academia and government. Through a variety of teams, Microsoft aims to empower these individuals and organisations - who we know are more targeted by threat actors - with the knowledge and tools to identify and prevent the spread of false or misleading information including deceptive content created using AI, such as deepfakes. We also equip eligible customers with our AccountGuard service free of charge to add an extra layer of cybersecurity protection, which has been offered globally since 2018. Political candidates or election authorities who have a concern about a deepfake of themselves can report it to our webpage where we can investigate and take action.
3. **Strengthening the information ecosystem:** Microsoft recognises that the creation of content is where synthetic content starts. As part of our Responsible AI approach, we have guardrails in place to ensure our AI systems are developed responsibly and in ways that warrant people's trust. We have guardrails to keep Bing Image Creator safe and prevent harmful use. We also ensure that, if citizens ask Bing election-related questions, they will be offered authoritative sources such as the Australian Electoral Commission to access voting information.

Finally, Microsoft offers AccountGuard, a security service designed to enhance protections for organisations involved in democratic processes and civil society. The service provides an extra layer of cyber threat monitoring and notification of nation-state actors targeting high-risk customers who enrol. AccountGuard is available in 40 countries, including Australia.

These examples represent the work of Microsoft's Elections & Societal Resilience team (formerly Democracy Forward Initiative), a program launched in 2018 to coordinate and track work undertaken across the company to protect and strengthen democratic institutions.

Microsoft

Microsoft's Elections & Societal Resilience team works closely with campaigns, electoral commissions and other entities around the world to support the integrity of elections.

Microsoft remains steadfast in its commitment to supporting the security and integrity of democratic processes. Through our comprehensive programs and collaborative efforts, we aim to protect democracy from the evolving threats posed by nation-state actors.

Microsoft

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 **Bing Search, MSN, Microsoft Advertising** and **LinkedIn** through the application of our internal policies and scalable measures.

(O.1a) Bing Search

Bing Search is an online search engine that provides a searchable index of websites available on the internet. Bing Search does not have a news feed for users, allow users to post and share content, or otherwise enable content to go "viral" on its service. Nonetheless, disinformation could at times appear in organic search results and we take active steps to counter it. As discussed above, 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.

Bing Search's primary mechanism for combatting misinformation in search is via ranking improvements that take into account the quality and credibility (**QC**) of a website and work to rank higher quality and more authoritative pages over lower authority content. Bing Search describes the main parameters of its ranking systems, including QC, in depth in How Bing Delivers Search Results. Abusive techniques and examples of prohibited SEO activities are described in more detail in the Bing Webmaster Guidelines.

Determining the QC of a website includes evaluating the clarity of purpose of the site, its usability, and presentation. QC also consists of an evaluation of the page's "authority", which includes factors such as:

- **Reputation**: What types of other websites link to the site? A well-known news site is considered to have a higher reputation than a brand-new blog.
- **Level of discourse:** Is the purpose of the content solely to cause harm to individuals or groups of people? For example, a site that promotes violence or resorts to name-calling or bullying will be considered to have a low level of discourse, and therefore lower authority, than a balanced news article.
- **Level of distortion**: How well does the site differentiate fact from opinion? A site that is clearly labelled as satire or parody will have more authority than one that tries to obscure its intent.
- **Origination and transparency of the ownership**: Is the site reporting first-hand information, or does it summarise or republish content from others? If the site doesn't publish original content, do they attribute the source?

In addition to its ranking algorithms, Bing Search's general abuse/spam policies prohibit certain practices intended to manipulate or deceive the Bing Search algorithms, including those that could be employed by malicious actors in the spread of disinformation. Bing's spam policies are detailed in the "Abuse and Examples of Things to Avoid" section of the Bing Webmaster Guidelines.

Although the Bing Search algorithm endeavours to prioritise relevance, quality, and credibility in all scenarios, in some cases Bing Search identifies a threat that undermines the efficacy of its algorithms. When this happens, Bing Search employs "defensive search" strategies and interventions to counteract threats.

Defensive search interventions may include:

- algorithmic interventions (such as authority signal boost in ranking or demotions of a website);
- restricting autosuggest or related search terms to avoid directing users to potentially problematic queries; and
- in limited cases, manual interventions for individual reported issues or broader areas more prone to misinformation or disinformation (e.g., elections, pharmaceutical drugs, or COVID-19).

Bing actively monitors manipulation trends in identified high-risk areas and deploys mitigation methods as needed to ensure users are provided with high quality, high authority search results.

Defensive Search Interventions, Australia

	January – December 2022		**January – December 2023**		**January – December 2024**		**April– December 2025***	
	Queries	**Impressions**	**Queries**	**Impressions**	**Queries**	**Impressions**	**Queries**	**Impressions**
Total	64,104	4,441,099	136,450	2,270,775	225,062	3,671,143	324,965	6,358,885
Ukraine related[#]	45,100	1,072,939	17,964	618,366	16,147	489,338	18,726	248,549

[#]Ukraine data from February to December 2022 for the 2022 reporting year.
*Queries and impressions data for 2025 is from April to December

Increased queries and impressions for Defensive Search Interventions in 2025 correlate to increased use of Bing in Australia. Queries and impressions related to Ukraine is the only category that is reported distinctly. The decrease in impressions means that users are less frequently selecting search results related to this topic.

Bing regularly partners with independent third-party organisations to obtain threat intelligence on emerging narratives and mis/disinformation patterns and tactics that helps to inform potential algorithmic interventions. Bing Search also takes action to remove auto-suggest and related search terms that could inadvertently result in problematic or misleading content. Bing Search also may include answers or public service announcements at the top of search results pointing users to high authority information on a searched topic or warnings on particular URLs known to contain harmful information (such as unaccredited online pharmacies and sites containing malware).

While Bing Search generally strives to rank its organic search results so that trusted, authoritative news and information appear first and provides tools that help Bing Search 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. Bing users have many legitimate reasons for seeking out content in search that may be harmful or offensive in other contexts (such as for research purposes) and unduly restricting access to information can pose risks to users' fundamental rights.

Generative AI Features

For AI-powered experiences, Bing has partnered closely with Microsoft's Responsible AI team to proactively address AI-related risks and continues to evolve these features based on user and external stakeholder feedback. Bing generative AI experiences continue to rely on the same infrastructure and mitigations previously discussed in Microsoft's last report.

Copilot Search's primary functionality is, like traditional Bing search, to provide users with links to third party content responsive to their search queries. As such, the ranking algorithms and spam/abuse policies described above continue to be Bing's primary defence against manipulation and abuse, supplemented by interventions designed to specifically address manipulation in generative AI features. Additional information on how Microsoft

approached responsible AI in Bing's generative AI experiences is available How Bing Delivers Search Results.

We note that Bing does not host user content and users cannot post or share content directly on the Bing service, including the generative AI experiences. In addition, Microsoft undertakes specific mitigations to address the risks that individuals may attempt to use generative AI to create deep fakes or manipulated media to spread misinformation. Although Bing does not have the ability to monitor third party platforms for publication of content created through Bing's services, Bing has implemented safeguards to help to minimise the risk that bad actors can use Bing generative AI experiences to create mis/disinformation that could potentially be shared on other platforms. See more here, here and here.

(O.1a) MSN

MSN delivers high-quality news across web and mobile experiences for Microsoft as well as a growing number of syndication partners. MSN'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 must adhere to a strict set of standards that prohibit false information, propaganda and deliberate misinformation.
- MSN is free to download, with no limits on number of articles or videos a user can view.

MSN saw a decline in proportion of misinformation takedowns in 2025

Misinformation remained a very small share of total MSN Australia comment takedowns, declining from 0.77% in 2024 to 0.34% in 2025. While the absolute number of misinformation-labelled takedowns increased year-on-year, the proportional share fell materially because overall takedowns increased significantly in 2025. The low share of misinformation-related comment takedowns can be attributed to several factors.

Firstly, takedown volumes for some of the specific misinformation narratives we track (e.g., COVID-19 and QAnon) declined compared to prior years, while others remained low overall.

Secondly, MSN continued to expand automated, LLM-enabled comment moderation, including initiating migration of portions of the comment moderation prompts from GPT-4o to GPT-5 during 2025. The data presented in these reports relates to user-initiated complaint and takedown processes; as more violating comments are proactively blocked or actioned by automated systems before users encounter them, there are fewer violative comments available for users to manually report.

MSN Community supports diverse, authentic conversations and content about issues and events. Our Community Guidelines 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.

When necessary, MSN 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.

We have specific policies for managing false or misleading information relating to well-defined narratives with potential for real-world harm - which includes disabling comments for certain articles to reduce propagation.

In the reporting period, 988,892 comments were proactively blocked in Australia by these systems on MSN.

MSN Comments, Australia Takedowns

	Oct–Dec 2021#		Jan–Dec 2022		Jan–Dec 2023		Jan–Dec 2024		Jan–Dec 2025	
Total takedowns	73,700	100%	849,000	100%	9,955	100%	8,464	100%	47,083	100%
Misinfo – all*	1,899	2.5%	9,256	1.09%	49	0.49%	65	0.77%	161	0.34%
Misinfo – COVID-19	1,810	2.4%	8,655	1.01%	33	0.33%	50	0.59%	10	0.02%
Misinfo – Qanon	89	0.12%	425	0.05%	4	0.04%	2	0.000%	0	0%
Misinfo - Russia/Ukraine^			128	0.01%	12	0.12%	4	0.000%	4	0.01%
Misinfo – US Elections							8	0.000%	66	0.14%

#Comments data prior to October 2021 is not a reliable metric as the function was only in its early stages.

* Misinformation total includes comments which have more than one trait labelled; percentages are rounded; sub-category list is not exhaustive.

^Russia/Ukraine misinformation trait was introduced in February 2022.

The misinformation categories shown in the table above represent well defined, high-harm narrative traits used for comment moderation. In parallel, MSN has expanded from a static

topic-only approach toward a more dynamic continuously refreshed misinformation strategy. This strategy ingests updated fact-check claims (e.g., from Reuters and AFP) on an ongoing basis and uses GPT-based classification to identify misinformation across markets enabling coverage of emerging narratives without requiring a new category or trait for every new misinformation event.

(O.1a) Microsoft Advertising

Microsoft Advertising's Information Integrity and Misleading Content Policies 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.

In 2025, Microsoft Advertising took action to ensure a safe and trusted experience.

This included:

- taking down more than ~8.8 billion ads and product offers for various policy violations. We suspended nearly 183,546 customers and blocked 512,442 ads that spread disinformation
- making use of significant advancements in 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;
- developing automated detection mechanisms to enforce new policies on information integrity, including developing new logic in the system to prevent receiving requests to show ads on web domains that may violate our disinformation policies; and,
- further iterating those automated detection mechanisms, including new automated classifiers to detect misleading claims relating to false information and consumer scams, such as financial scams, unsupported pricing claims and sensationalised ads, and misleading celebrity endorsements.

Microsoft Advertising deploys a range of policy-based and proactive measures to reduce the risk of harms associated with disinformation and misinformation, including:

- Our Relevance and Quality Policies, which 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).

- Our Critical Events Policy, 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 affect user safety.

Microsoft Advertising maintains and enforces network-wide policies designed to prevent the publishing and carriage of harmful disinformation and the placement of advertising next to disinformation content. Such policies prohibit ads or sites that contain or lead to disinformation. Our policy states, "We may use a combination of internal signals and trusted third-party data or information sources to reject, block, or take down ads or sites that contain disinformation or send traffic to pages containing disinformation. We may block at the domain level landing pages or sites that violate this policy." See our main policy page.

In 2025, Microsoft Advertising enhanced its detection capabilities through continuous integration with services provided by the Microsoft Threat Analysis Center (MTAC), including signals related to Foreign Information Manipulation and Interference (FIMI) domains. MTAC operates as part of Microsoft's broader threat intelligence structure alongside the Microsoft Threat Intelligence Center (MSTIC) and the Digital Crimes Unit and focuses on identifying foreign malign influence operations and state-aligned information campaigns. MTAC applies a geopolitical influence-operations lens to analyse propaganda, coordinated information operations, and state-backed manipulation campaigns, including cross-platform narrative efforts and manipulated or deceptive content targeting public opinion.

Microsoft Advertising regularly consumes domain- and web-property-level intelligence informed by MTAC analysis to detect ads associated with disinformation, misinformation, impersonation, or influence operations.

As previously reported, Microsoft Advertising is continuing to prevent serving advertising related to the Russia-Ukraine conflict, pursuant to its Critical Events Policy. Relatedly, Microsoft Advertising is preventing serving advertising related to the Israel-Hamas conflict pursuant to its Critical Events Policy. Under this policy, Microsoft Advertising reserves the right to remove or limit advertising in response to a sensitive or high-profile news event to prevent the commercial exploitation of such events and to ensure user safety.

Microsoft Advertising Global Ad Takedowns

	2021	2022	2023	2024	2025
Global ad takedowns	3 billion	7.2 billion	8.2 billion	7.9 billion	8.8 billion

Microsoft Advertising Ad Safety in Australia

	2022		2023		2024		2025	
Action	**Global**	**Australia**	**Global**	**Australia**	**Global**	**Australia**	**Global**	**Australia**
Rejections	7.2b	1b	8.2b	1.33b	7.9b	1.45b	8.8 b	1.7 b
Total appeals	127,158	14,536	132,910	7,747	2,199,275	43,252	2,977,361	77,307
Total appeals overturned	101,537	9,522	95,738	6,858	1,563,677	32,409	2,446,125	61,416
Total complaints	35,667	285	46,168	1,090	48,470	425	50,622	1,277
Complaint: Policy violation	1,156	57	1,411	14	477	6	9,875	823
Complaint: Trademark infringement	32,213	153	31,223	238	26,037	223	39,512	434
Complaint: User safety issues	1,805	58	13,533	838	20,352	177	796	20
Complaints : Other	493	17	407	7	1,599	19	439	0
Total entity takedowns	551,424	118,321	1,746,324	332,332	1,201,698	1,195,323	727,137	722,299
Average processing time	~36 hours	~36 hours	~36 hours	~36 hours	~36 hours	~36 hours	~36 hours	~36 hours

The proximity between the Australian and global figures for total entity takedowns reflects Microsoft Advertising's updated global enforcement model, where actions such as ad rejections and removals are applied at a global level, resulting in broadly similar figures across jurisdictions. Minor differences between jurisdiction-specific and global figures may still arise due to residual workflow and reporting variations, including the treatment of manual takedowns versus automated rejections, jurisdiction-specific targeting of enforcement actions, and differences in reporting scope or data extraction timing across frameworks.

Year-on-year changes are driven by a combination of factors, including a shift toward proactive enforcement (blocking ads before they serve), as well as changes in methodology,

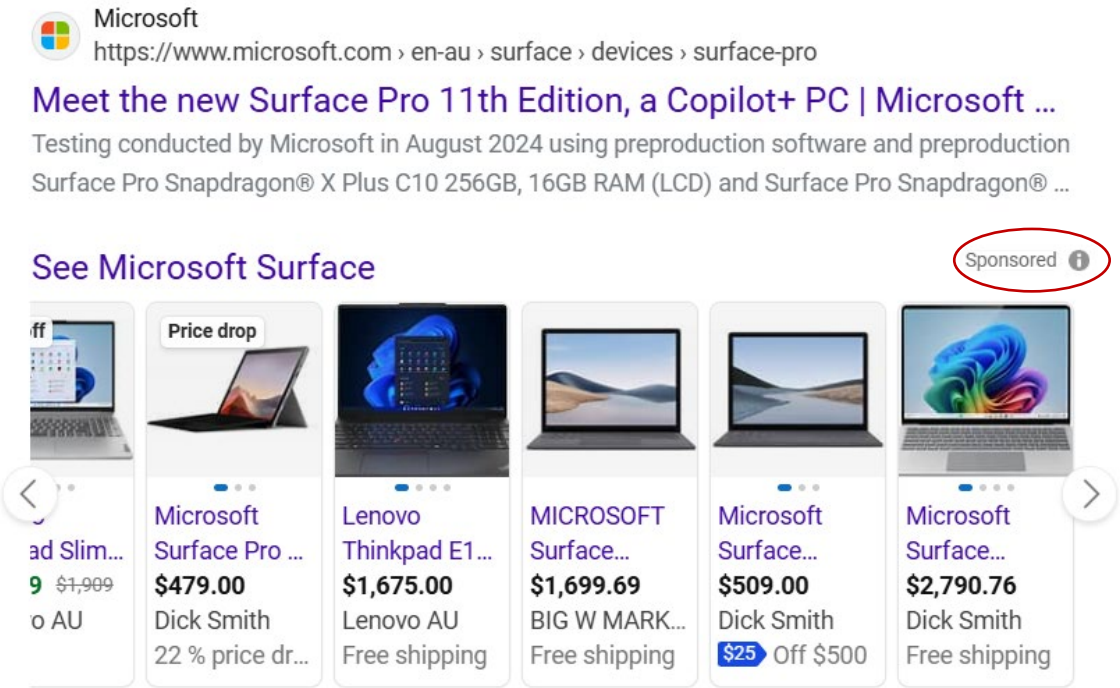
such as moving from a mix of suppression and rejection to primarily global rejection-based metrics and updated definitions of removal and prohibited ads.

Microsoft's Advertiser Identify Verification Program

The Advertiser Identity Verification Program, designed to verify the identity of the advertisers who buy ads through Microsoft Advertising, is available across our ad network, including in Australia. In 2025, 7,913 accounts opted for AIV verification in Australia and out of these, 6,451 accounts were successfully verified. The system enables customers to 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.

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

All Microsoft services that display ads served by Microsoft Advertising clearly distinguish sponsored from non-sponsored content by displaying a 'Sponsored' 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 ad setting page.



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. Microsoft Advertising proactively reviews publisher partners to enforce this requirement.

(O.1a) LinkedIn

To help keep **LinkedIn** safe, trusted, and professional, LinkedIn's Professional Community Policies 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 remove inauthentic activity. We're regularly rolling out scalable technologies like machine learning models to keep our platform safe.

Using the process described in response to Outcome 1c 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 actioned in accordance with our policies.

LinkedIn has numerous workstreams that address misinformation, particularly during crisis situations. For instance, LinkedIn's in-house editorial team provides members with trustworthy content regarding global events, including Russia's war against Ukraine and the Israel-Hamas conflict. LinkedIn has an internal team of hundreds of content reviewers located all over the world providing 24/7 coverage and includes specialists in a number of languages.

LinkedIn's 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.

LinkedIn Community Report: global actions taken on content that violated Professional Community Policies and User Agreement, January 2021 – December 2025

	2021 Jan-Jun	**2021 Jul-Dec**	**2022 Jan-Jun**	**2022 Jul-Dec**	**2023 Jan-Jun**	**2023 Jul-Dec**	**2024 Jan-Jun**	**2024 Jul-Dec**	**2025 Jan-Jun**	**2025 Jul-Dec**
Fake Accounts Stopped at registration	11.6m	11.9m	16.4m	44.7m	42.5m	46.3m	70.1m	80.6m	61.2m	88.9m
Restricted proactively	3.7m	4.4m	5.4m	13.2m	15.1m	17.1m	16.2m	19.7m	22.2m	24.4m

	2021 Jan-Jun	2021 Jul-Dec	2022 Jan-Jun	2022 Jul-Dec	2023 Jan-Jun	2023 Jul-Dec	2024 Jan-Jun	2024 Jul-Dec	2025 Jan-Jun	2025 Jul-Dec
Restricted after report	85.7k	127k	190k	201k	196k	232.4k	262.9k	265.7k	385.9k	361.3k
Content Violation Mis-information	147.5k	207.5k	172.4k	138k	85.2k	53.8k	30.5k	62.2k	36.2k	8.4k

LinkedIn Community Report: actions taken in Australia on content that violated Professional Community Policies and User Agreement, January 2021 – December 2025

	2021 Jan-Jun	2021 Jul-Dec	2022 Jan-Jun	2022 Jul-Dec	2023 Jan - Jun	2023 Jul - Dec	2024 Jan - Jun	2024 Jul - Dec	2025 Jan - Jun	2025 Jul - Dec
Fake Accounts Stopped at registration	54,883	45,983	81,533	149,591	112,767	253,569	399,817	347,953	360,721	783,615
Restricted proactively	64,642	39,179	63,317	112,809	116,613	126,502	142,134	161,970	155,224	165,905
Restricted after report	1,281	1,448	1,755	2,023	2,168	2,633	1,374	1,074	1,919	4,404
Content Violation Mis-information*	2,149	6,007	3,946	1,656	969	571	456	804	633	300
Mis-information content removals that were appealed by the content author		219	151	79	13	17	18	11	32	33
The number of appeals that were granted		3	3	3	1	3	0	0	10	9

+ Since July – December 2022, LinkedIn stopped more fake accounts compared to previous periods. Because of our multidimensional approach to combating fake accounts, the manner in which we detect fake accounts changed a bit from July 2022 onwards. With the rise of fraudulent activity taking place across the internet,

LinkedIn continues to treat fake accounts as a top priority and invest in additional verification features, safety tools, and automated defences to support safe experiences.

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
- **MSN:** Microsoft Services Agreement, Community Guidelines
- **LinkedIn:** User Agreement, Professional Community Policies

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

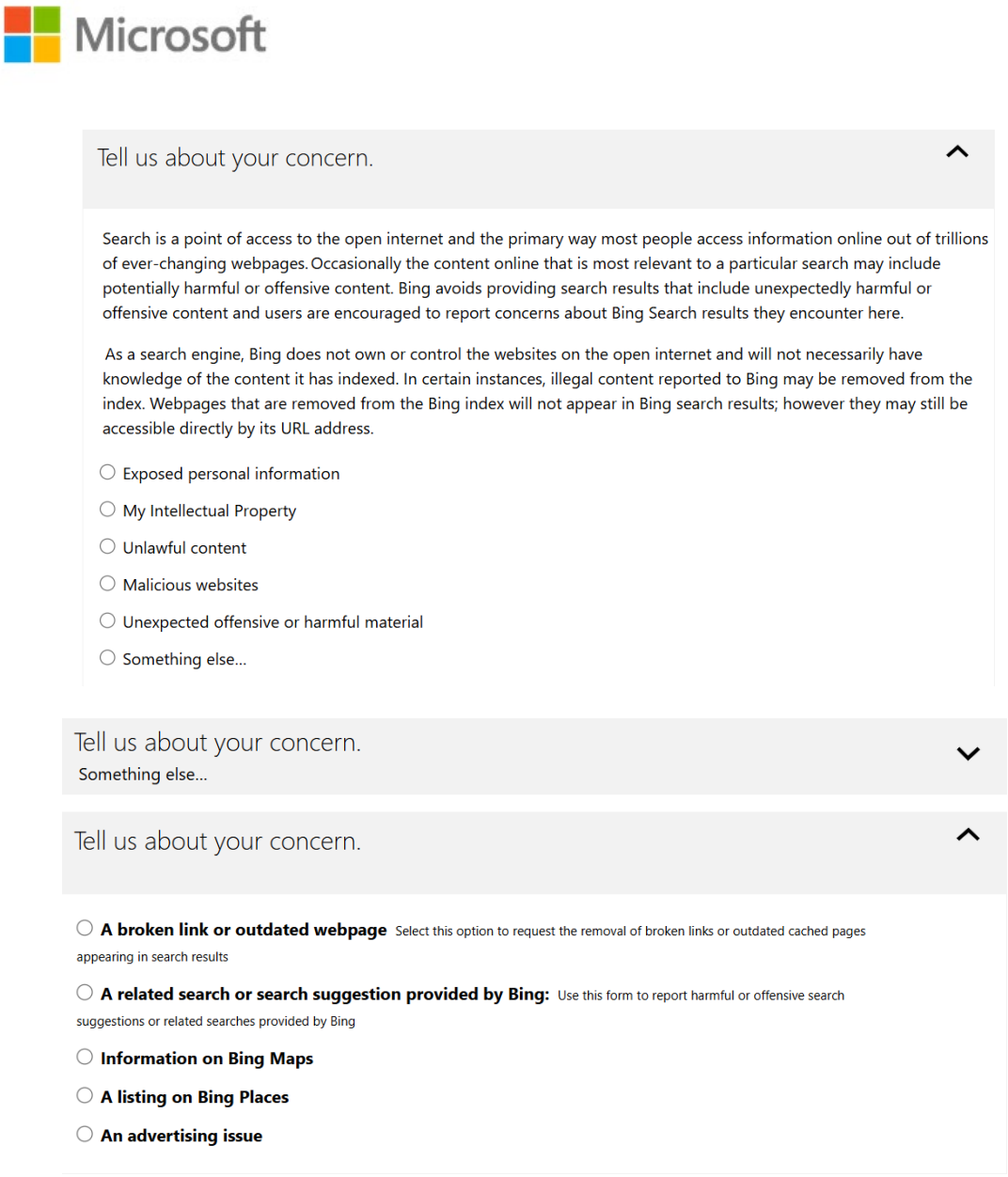
In addition to the guidelines contained within the respective user agreements, **Bing Search, MSN, Microsoft Advertising** and **LinkedIn** have reporting mechanisms where users are able to flag problematic content.

(O.1c) Bing Search

Bing Search has updated its "Report a Concern" and "Feedback" tools to include enhanced reporting for generative AI features as well as traditional web search. Bing Search's Report a Concern Form permits users to report third-party websites for a variety of reasons including disclosure of private information, spam and malicious pages, and illegal materials. Bing Search's "Feedback" tool, which is accessible on the lower right corner on a search results page, allows users to provide feedback on search results (including a screenshot of the results page) to Bing Search.

These tools have also been updated to make it easy for users to report problematic content they encounter while using Copilot in Bing by including the same "Feedback" button with direct links to the respective service's "Report a Concern" tool on the footer of each page of Copilot.

Depending on the nature of the feedback, Bing Search may take appropriate action, such as to engage in algorithmic interventions to ensure high authority content appears above low authority content in search results, remove links that violate local law or Bing policies, add answers, warnings or other media literacy interventions on certain topics, or remove auto-suggest terms.



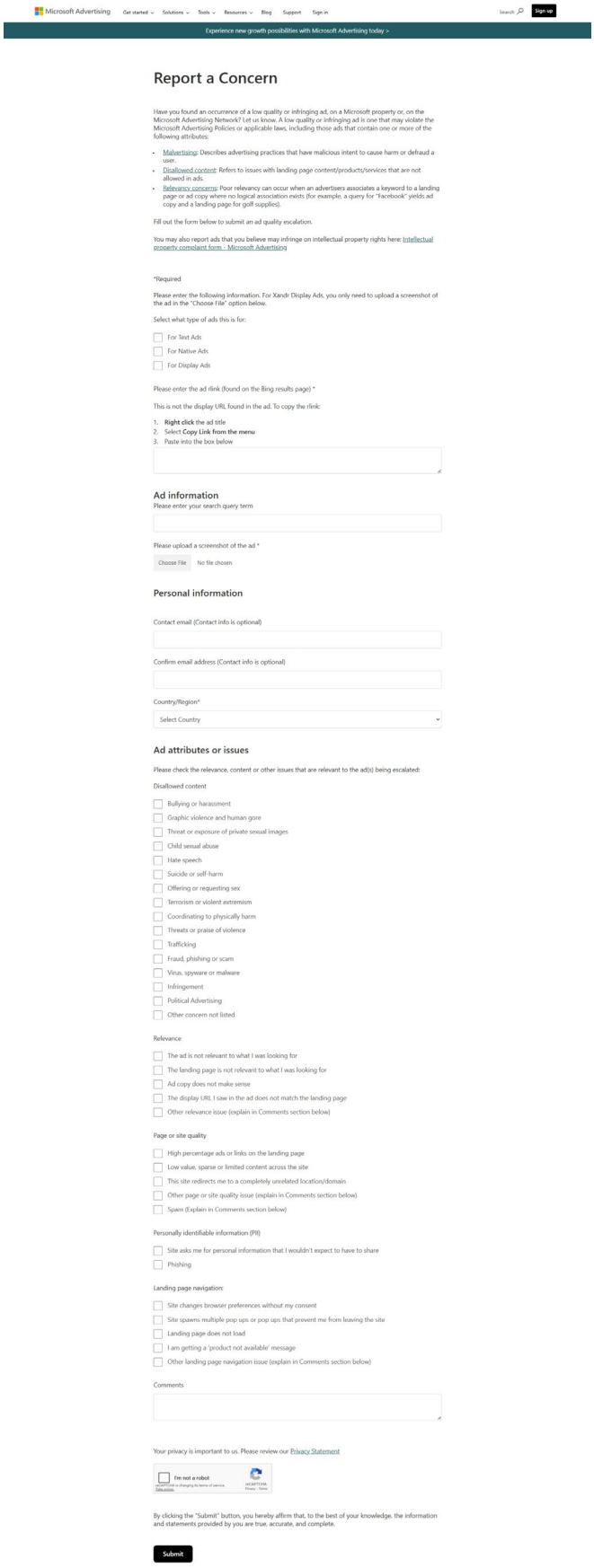
(O.1c) MSN

MSN 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. In addition, each article includes a 'Report an issue' option, with 'misleading title', 'outdated article' and 'suspected AI/bot created' available as types of issues.

(O.1c) 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 'Report a Concern' (as shown below). Users of Bing Search and MSN can also report ads via the respective feedback functions on those services.

Microsoft



Microsoft Advertising Get started Solutions Tools Resources Blog Support Sign in Search Sign up

Experience new growth possibilities with Microsoft Advertising today >

Report a Concern

Have you found an occurrence of a low quality or infringing ad, on a Microsoft property or, on the Microsoft Advertising Network? Let us know. A low quality or infringing ad is one that may violate the Microsoft Advertising Policies or applicable laws, including those ads that contain one or more of the following attributes:

- Malvertising: Describes advertising practices that have malicious intent to cause harm or defraud a user.
- Disallowed content: Refers to issues with landing page content/products/services that are not allowed in ads.
- Relevancy concerns: Poor relevancy can occur when an advertisers associates a keyword to a landing page or ad copy where no logical association exists (for example, a query for "Facebook" yields ad copy and a landing page for golf supplies).

Fill out the form below to submit an ad quality escalation.

You may also report ads that you believe may infringe on intellectual property rights here: Intellectual property complaint form - Microsoft Advertising

*Required

Please enter the following information. For Xandr Display Ads, you only need to upload a screenshot of the ad in the "Choose File" option below.

Select what type of ads this is for:

- [ ] For Text Ads
- [ ] For Native Ads
- [ ] For Display Ads

Please enter the ad rlink (found on the Bing results page) *

This is not the display URL found in the ad. To copy the rlink:

1. **Right click** the ad title
2. Select **Copy Link from the menu**
3. Paste into the box below

Ad information

Please enter your search query term

Please upload a screenshot of the ad *

Choose File No file chosen

Personal information

Contact email (Contact info is optional)

Confirm email address (Contact info is optional)

Country/Region*

Select Country

Ad attributes or issues

Please check the relevance, content or other issues that are relevant to the ad(s) being escalated:

Disallowed content

- [ ] Bullying or harassment
- [ ] Graphic violence and human gore
- [ ] Threat or exposure of private sexual images
- [ ] Child sexual abuse
- [ ] Hate speech
- [ ] Suicide or self-harm
- [ ] Offering or requesting sex
- [ ] Terrorism or violent extremism
- [ ] Coordinating to physically harm
- [ ] Threats or praise of violence
- [ ] Trafficking
- [ ] Fraud, phishing or scam
- [ ] Virus, spyware or malware
- [ ] Infringement
- [ ] Political Advertising
- [ ] Other concern not listed

Relevance

- [ ] The ad is not relevant to what I was looking for
- [ ] The landing page is not relevant to what I was looking for
- [ ] Ad copy does not make sense
- [ ] The display URL I saw in the ad does not match the landing page
- [ ] Other relevance issue (explain in Comments section below)

Page or site quality

- [ ] High percentage ads or links on the landing page
- [ ] Low value, sparse or limited content across the site
- [ ] This site redirects me to a completely unrelated location/domain
- [ ] Other page or site quality issue (explain in Comments section below)
- [ ] Spam (Explain in Comments section below)

Personally identifiable information (PII)

- [ ] Site asks me for personal information that I wouldn't expect to have to share
- [ ] Phishing

Landing page navigation:

- [ ] Site changes browser preferences without my consent
- [ ] Site spawns multiple pop ups or pop ups that prevent me from leaving the site
- [ ] Landing page does not load
- [ ] I am getting a 'product not available' message
- [ ] Other landing page navigation issue (explain in Comments section below)

Comments

Your privacy is important to us. Please review our Privacy Statement

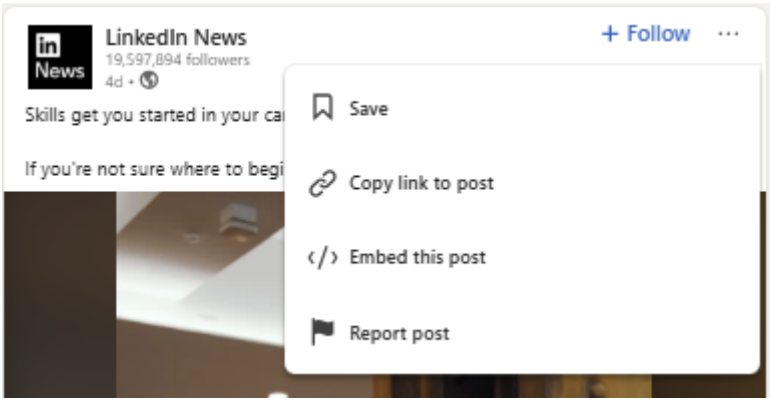
I'm not a robot reCAPTCHA

By clicking the "Submit" button, you hereby affirm that, to the best of your knowledge, the information and statements provided by you are true, accurate, and complete.

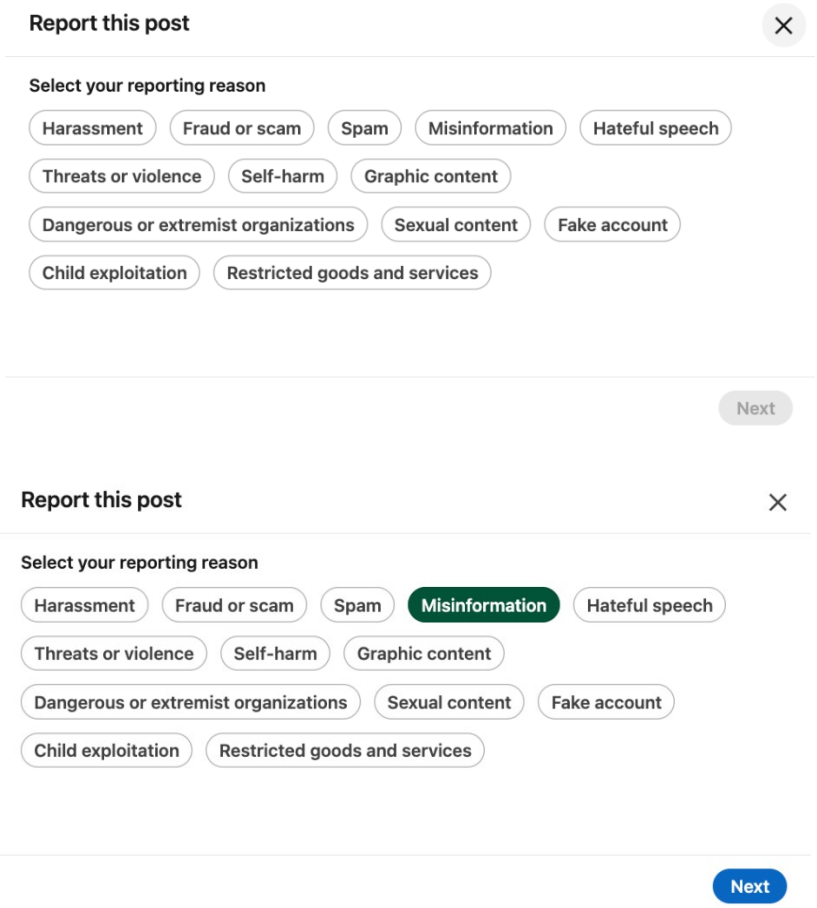
Submit

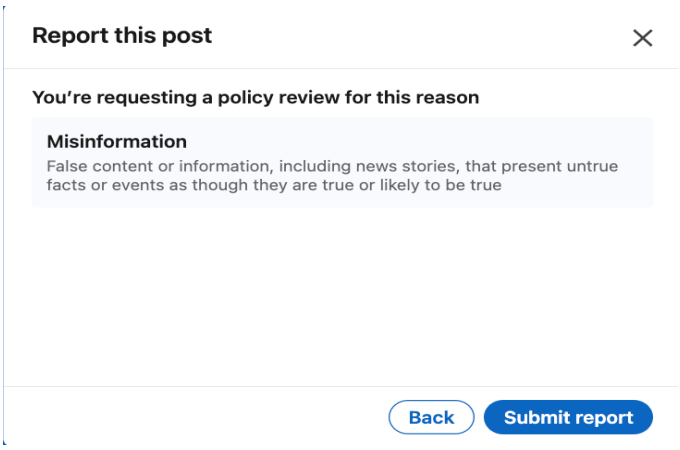
(O.1c) LinkedIn

If **LinkedIn's** members identify content they believe violates our Professional Community Policies, we encourage them to report it using the in-product reporting mechanism represented by the three dots in the upper right-hand corner of the content itself on LinkedIn or the "More" button on the Profile headline:



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





Reported content is generally enqueued for human review by LinkedIn's Trust Review Operations team or by LinkedIn's automated systems. If reported content is found to violate the Professional Community Policies, it will be actioned in accordance with our policies.

Both reporter and creator generally are notified when content is removed and both are given an opportunity to appeal the decision. Notices are typically sent by email and contain a link to a notice page containing certain additional information (e.g., about the content at issue, the policy violated, the action LinkedIn has taken and, in most instances, a link to allow the individual to appeal LinkedIn's decision). LinkedIn reviews those appeals and notifies the member of its appeal decision.

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

(O.1d) Bing Search, MSN, Microsoft Advertising

In addition to the sources detailed below, 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.

(O.1d) LinkedIn

As noted in our response to O.1a above, **LinkedIn's** 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.

Outcome 1e: Users will be able to access general information about Signatories' use of recommender systems and have options relating to content suggested by recommender systems.

(O.1e) LinkedIn

LinkedIn has published, and continues to publish, a variety of articles and other content, to explain to users how our recommender systems work, and how members can tailor their experiences on the LinkedIn platform, including:

- LinkedIn: Where Real Conversations Matter
- How Does the LinkedIn Feed Work?
- The Tools & Insights To Help You Tailor Your LinkedIn Feed Experience
- Understanding how the LinkedIn Feed Works
- Mythbusting the Feed: How the Algorithm Works
- Mythbusting the Feed: Helping our members better understand LinkedIn
- Our approach to building transparent and explainable AI systems
- Guide: Features to Help You Control Your Feed and Conversations

LinkedIn also makes easily accessible in the footer of every LinkedIn page a link out for "Recommendation Transparency", which links to a Help Center article about how LinkedIn ranks content for the member. That article links to more information about how members can customise their feed, including the ability for members to sort their most relevant posts chronologically in desktop view.

Additionally, LinkedIn addresses automated processing and relevancy in the LinkedIn User Agreement at the end of Section 3.6 and in our LinkedIn relevance - Optimizing the member experience Help Center article.

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. Demonetisation is one of Microsoft's core Information Integrity Principles, which outlines how we will not wilfully profit from foreign cyber influence content or actors. 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.

(O.2) Microsoft Advertising

In December 2022, **Microsoft Advertising** rolled out revised network-wide policies to avoid the publishing and carriage of harmful disinformation and the placement of advertising next to disinformation content. Such policies prohibit ads or sites that contain or lead to disinformation. To enforce this policy, we may use a combination of internal signals and trusted third-party data or information sources to reject, block, or take down ads or sites that contain disinformation or send traffic to pages containing disinformation. We may block at the domain level landing pages or sites that violate this policy.

In 2025, Microsoft Advertising enhanced its detection capabilities through continuous integration with services provided by the Microsoft Threat Analysis Center (MTAC), including signals related to Foreign Information Manipulation and Interference (FIMI) domains. MTAC operates as part of Microsoft's broader threat intelligence structure, alongside the Microsoft Threat Intelligence Center r(STIC) and the Digital Crimes Unit and focuses on identifying foreign malign influence operations and state-aligned information campaigns. MTAC applies a geopolitical influence-operations lens to analyse propaganda, coordinated information operations, and state-backed manipulation campaigns, including cross-platform narrative efforts and manipulated or deceptive content targeting public opinion.

Microsoft Advertising regularly consumes domain- and web-property-level intelligence informed by MTAC analysis to detect ads and publishers associated with disinformation, misinformation, impersonation, or influence operations.

Additionally, Microsoft Advertising's policies with respect to publishers include a comprehensive list of prohibited content that ads cannot serve against. Prohibited content includes, but is not limited to:

- disinformation;
- sensitive content (e.g., extreme, aggressive, or misleading interpretations of news, events, or individuals);
- unmoderated user-generated content; and

- unsavoury content (such as content disparaging individuals or organisations).

Publishers are required to maintain a list of prohibited terms and provide us with information on their content management practices where applicable. In addition to content requirements, publishers are required to abide by restrictions against engaging in business practices that are harmful to users (e.g., distributing malware).

Advertisers who willingly or repeatedly violate our terms or policies are suspended from accessing the service and cannot service ads until they redress the violation.

(O.2) LinkedIn

LinkedIn prohibits misinformation and disinformation on its platform, whether in the form of organic content or in the form of advertising content.

LinkedIn's Professional Community Policies, which apply to all content on LinkedIn's platform expressly prohibit false and misleading content, including misinformation and disinformation. LinkedIn provides additional specific examples of false and misleading content that violates its policy via a Help Center article on False or Misleading Content.

LinkedIn's Advertising Policies incorporate the above provision, and similarly prohibit misinformation and disinformation. In addition, LinkedIn's Advertising Policies also prohibit fraudulent and deceptive ads and require any claims made in an ad have factual support.

LinkedIn members may also report ads that they believe violate LinkedIn's advertising policies and, when members report ads, LinkedIn's Advertising Review team reviews them. To report an ad, members can click on the three-dot icon in the upper right-hand corner of every ad and select the "Hide or report this ad" option.

LinkedIn provides a range of information and tools to give advertisers transparency and control regarding the placement of their advertising. For example, for ads on the LinkedIn platform, LinkedIn publishes a Feed Brand Safety score for advertisers and the public. The Feed Brand Safety score measures the number of ad impressions on the LinkedIn platform that appeared adjacent to – that is, immediately above or below within the LinkedIn feed – content removed for violating LinkedIn's Professional Community Policies, including disinformation. From July through December 2025, the Feed Brand Safety score was 99%+ safe. More information about LinkedIn's Feed Brand Safety Score.

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), **Bing Search, Microsoft Advertising**, and **LinkedIn** reduce the risk of inauthentic user behaviours through the measures detailed below.

(O.3) Bing Search

The "Abuse and Examples of Things to Avoid" section of the Bing Webmaster Guidelines details the policies intended to maintain the integrity of Bing Search. Bing's general spam policies prohibit certain practices intended to manipulate or deceive the Bing search algorithms.

Bing may take action on websites employing spam tactics or that otherwise violate the Webmaster Guidelines, including by applying ranking penalties (such as demoting a website or delisting a website from the index). However, it is important to clarify that in search it is not feasible to distinguish between spam tactics employed by malicious actors specifically for the purpose of spreading disinformation and other types of spam.

In addition to enforcing its spam policies, Bing takes actions to promote high authority, high quality content and thereby reduce the impact of disinformation appearing in Bing search results. Among other initiatives, this includes:

- continued improvement of its ranking algorithms to ensure that the most authoritative, relevant content is returned at the top of search results;
- regular review and actioning of disinformation threat intelligence;
- contributing to and supporting the research community; and
- implementation and enforcement of clear policies concerning the use of manipulative tactics on Bing Search.

Although the Bing search algorithms endeavour to prioritise relevance, quality, and credibility in all scenarios, in some cases Bing identifies a threat that undermines the efficacy of its algorithms. When this happens, Bing employs "defensive search" strategies and interventions to counteract threats in accordance with its trustworthy search principles to help protect Bing users from being misled by untrustworthy search results and/or inadvertently being exposed to unexpected harmful or offensive content.
In addition to defensive search, Bing Search regularly monitors for violations of its Webmaster Guidelines, including attempts to manipulate the Bing search algorithms through prohibited practices such as cloaking, link spamming, keyword stuffing, and phishing.

(O.3) 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.

(O.3) LinkedIn

LinkedIn's professional focus shapes the type of content we see on platform. People tend to say things differently when their colleagues and employer are watching. Accordingly, our members do not tend to use LinkedIn to engage in the mass dissemination of misinformation, and bad actors generally need to create fake accounts to peddle misinformation.

To ensure their content reaches a large audience, bad actors need to either connect with real members or post content that real members will like— both of which are hard to achieve on LinkedIn given our professional focus. The mass dissemination of false information, as well as artificial traffic and engagement, therefore, requires the mass creation of fake accounts, which we have various defences to prevent and limit.

To respond to the ever-changing threat landscape, LinkedIn's teams continually invest in new technologies for combating inauthentic behaviour on the platform. For example, LinkedIn is 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, such as deep fakes; anomaly detection of risky behaviours; and deep learning models for detecting sequences of activity that are associated with abusive automation.

LinkedIn has also adopted the Coalition for Content Provenance and Authenticity's industry-leading "Content Credentials" technology (C2PA) to include metadata labelling, including data about whether content is created using AI, on content containing the C2PA technology. Users will see the "Cr" label. By clicking the label, users will be able to trace the origin of the AI-created media, including the source and history of the content, and whether it was created or edited by AI.

LinkedIn acts vigilantly to maintain the integrity of all accounts and to ward off bot and false account activity.

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 our User Agreement prohibiting the use of "bots or other unauthorised automated methods to access the Services, add or download contacts, send or redirect messages, create, comment on, like, share, or re-share posts, or otherwise drive inauthentic engagement" through:

- Monitoring platform conversations regarding significant elections and establishing metrics for when election-related conversations, violations, or operational capacity breach a threshold and require additional support;
- Maintaining a dedicated Anti-Abuse team to research emerging trends and key risks and develop tools to address them;
- Using AI to detect inauthentic activity and communities of fake accounts;
- Conducting hash matching for known instances of deepfake content;
- Using automated systems to detect and block automated activity; and
- Maintaining 24/7 escalation paths to address any emerging issues.

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 users 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.

(O.4) Bing Search

Bing Search offers a number of tools to help users understand the context and trustworthiness of search results. Even in circumstances where a user is expressly seeking low authority content (or if there is a data void so little to no high authority content exists for a query), Bing Search provides tools to users that can help improve their digital literacy and avoid harms resulting from engaging with misleading or inaccurate content.

Bing Search is enhanced with various features to help users navigate complex information environments with confidence. These include, for example:

- Bing Search Intelligent Answers also provides users with informative panels and direct answers to certain search queries and is now available in 100 languages.
- Bing Search's "Knowledge Cards" feature gives users a single view of authoritative information on a specific topic and are typically displayed at the top of the SERP page.
- Bing Search's Page Insights feature also helps provide users with information and context about websites contained in the search results. The feature, which appears as a light bulb image next to certain search results, provides users with additional information about the site and its contents from third party information sites such as Wikipedia.
- Bing Search ingests tags for fact-check articles using the ClaimReview open schema to help users find fact checking information and warns users with red "flags" when fact-checked claims or content appearing in search results has been determined to be false or unfounded by third-party fact checkers;
- Microsoft continues to offer Search Coach as a free app in Microsoft Teams to help educators and students form effective queries and identify reliable resources. It is

designed to teach information literacy skills in a safe, secure, and ad-free environment.

Other initiatives Microsoft has engaged in to build digital literacy include:

- We piloted a global training for content creators and influencers in Australia in May 2025. The training reflects on the role of AI in the context of Australian elections while building creators' understanding of AI ethics and deceptive uses of AI so that they can make more informed decisions about their content and better educate their audiences.

- Ongoing partnerships with third-party organisations, including Microsoft and The Trust Project, to fund media literacy campaigns while continuing introductory calls with new organisations to grow additional campaigns' reach to new markets.

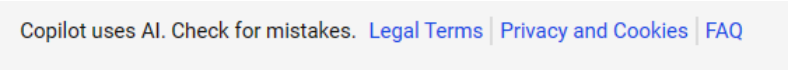
- Continued to provide pro-bono advertising space across Microsoft surfaces to disseminate the literacy campaigns and helped garner millions of impressions per month.

- Helped educators build AI literacy and make the most of AI capabilities, we offer a free module on Microsoft Learn: Enhancing teaching and learning with Copilot. This module is designed to guide educators through available features, learn how to create and iterate on prompts, and use expertise to evaluate responses for quality and credibility; and

- We offer *The Investigators*, a new world for Minecraft Education that helps students build information and media literacy through game-based learning. We have also launched *AI Foundations*, a suite of video, game and learning resources designed to build AI and information literacy.

Copilot

In addition to the features available for core search experiences, Copilot also provides information to help educate users on the uses and limitations of generative AI-driven search experiences, such as by reminding users that they are interacting with a generative-AI system and that mistakes can occur (see below):

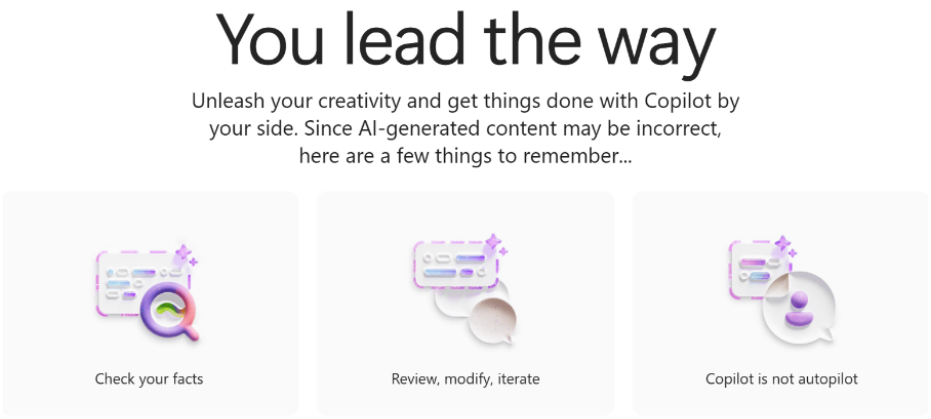


Explanatory documents like Copilot in Bing: Our Approach to Responsible AI also help educate users on the nature of AI-driven search experiences and the uses, safeguards, and limitations of this emerging technology.

For example, the Copilot FAQ answer to "Are Copilot's AI-generated responses always factual?" explains: "Copilot aims to base all its responses on reliable sources - but AI can make mistakes, and third-party content on the internet may not always be accurate or reliable. Copilot will sometimes misrepresent the information it finds, and you may see responses that sound convincing but are incomplete, inaccurate, or inappropriate. Use your

own judgment and double check the facts before making decisions or taking action based on Copilot's responses."

In another prominent example, another Copilot support page reminds users prominently that while they "lead the way", they need to check facts, review, and avoid simply relying on AI as an "autopilot" (shown below). Additionally, Microsoft has released a Classroom Toolkit for teachers that encourages responsible education around generative AI tools including the importance of fact checking.



Microsoft also offers meaningful resources for users interested in learning more about generative AI features and tools, including Copilot, though blog posts, articles, information hubs, and support pages. In addition to teaching AI basics and how-tos, these resources reiterate the importance of checking AI-generated materials and understanding the strengths and limitations of AI. See e.g., Microsoft 365 Copilot help & learning.

Microsoft is committed to providing resources, educational materials, and guides so that users can develop literacy when interacting with AI systems and will continue to explore ways to further educate the public on important generative AI topics.

(O.4) MSN

MSN clearly labels the sources of news articles and distinguishes advertising to enable users to readily differentiate this from other content.

(O.4) LinkedIn

As the world around us changes, **LinkedIn** continues to evolve and adapt our systems and practices for combating misinformation and other inauthentic behaviour on our platform, including to respond to the unique challenges presented by world events.

In addition to broader measures, LinkedIn has taken steps to tackle disinformation in connection with unfolding world events. LinkedIn's in-house editorial team provides members with trustworthy content regarding global events.

Further, LinkedIn's voluntary verification features allow members to verify certain information about the Pages they administer, the Jobs they post, or their Profile (like the member's

association with a particular company or educational institution, or their identity, using a government-issued ID through one of LinkedIn's identity verification partners). Verified information, which is marked by a badge symbol in various locations on the platform, provides members with authenticity signals about the people, Jobs and Pages they encounter on LinkedIn. These signals are designed to help members make more informed decisions about connecting or engaging with other professionals, organizations, or Jobs.

Currently, LinkedIn supports several forms of verification. Each form of verification has its own eligibility and availability criteria. Verifications are free, aimed at increasing trust and authenticity on the platform, and can be removed at any time by the member or entity whose information is verified.

(O.4) Other contributions and measures

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

Microsoft's commitments and actions under the Tech Accord

Microsoft and LinkedIn remain committed to the Tech Accord to Combat Deceptive Use of AI in 2024 Elections. Microsoft has continued to meet our commitments in the Tech Accord by implementing content provenance, establishing reporting channels and improving detection capability. For example:

- Microsoft has added Content Credentials to all images created with our most popular consumer facing AI image generation tools, including Bing Image Creator, Microsoft Designer, Copilot, and in our enterprise AI image generation tools including Azure OpenAI DALL-E. In addition, images with Content Credentials uploaded to LinkedIn are automatically labelled and the credential information is shown to end users.
- We have joined forces with fellow Tech Accord signatory Truepic and have jointly released an app that simplifies the process of adding Content Credentials to authentically captured smartphone camera images, video and audio for participants of the pilot.
- We are also working with our partners at True Media to provide governments, civil society and journalists with access to free tools that enable them to check whether an image or video was AI generated and/or manipulated.
- Microsoft is harnessing the data science and technical capabilities of our AI for Good Lab and Microsoft Threat Analysis Center teams to better assess whether content – including that created and disseminated by foreign actors – is synthetic of not. Microsoft's AI for Good lab has developed and is using detection models (image, video) to assess whether media is generated or manipulated by AI. In February 2024, we launched the Microsoft-2024 Elections site. This site empowers candidates, campaigns or election authorities to directly report deceptive AI election content on Microsoft consumer services. This reporting tool allows for 24/7 reporting by impacted election entities who have been targeted by deceptive AI found on Microsoft platforms.
- As detailed above in 'Microsoft's role in the 2025 Australian Election' above, we continue to work closely with the Australian Electoral Commission along with political and media stakeholders in Australia to provide pre-election support. We have also continued our global efforts to conduct training sessions for political stakeholders.

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.

(O.5) Microsoft Advertising

Under our Advertising Policies, **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" are also prohibited.

(O.5) LinkedIn

LinkedIn does not accept political advertising. LinkedIn's Advertising Policies globally prohibit political ads, including but not limited to ads that:

- 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 PACs or similar organisations; and

- 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.

Microsoft

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.

A non-exhaustive list of Microsoft's ongoing collaborations with the broader research community in this space include:

Bing Search ORCAS dataset	Bing Search provides researchers with access to ORCAS: Open Resource for Click Analysis in Search a click-based dataset associated with the TREC Deep Learning Track, which provides 18 million connections to 10 million distinct queries and is available to researchers.
Responsible AI Toolbox	As a leader in research in Responsible AI, Microsoft provides a range of tools and resources dedicated to promoting responsible usage of AI to allow practitioners and researchers to maximise the benefits of AI systems while mitigating harms. For example, as part of its Responsible AI Toolbox, Microsoft provides a Responsible AI Mitigations Library, which enables practitioners to more easily experiment with different techniques for addressing failure (which could include inaccurate outputs), and the Responsible AI Tracker, which uses visualisations to show the effectiveness of the different techniques for more informed decision-making. These tools are available to the public and research community for free.
Partnership on AI	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). Microsoft funding has supported the development of 19 in-depth case studies on different companies' (including Microsoft's) experiences implementing PAI's *Responsible Practices for Synthetic Media* Framework including 8 case studies that explore audience understanding of AI transparency and disclosure practices.
MS MARCO	Bing Search makes information available to the research community to improve search results by making data sets like its MS MARCO publicly available. Bing Search provides researchers and the public with access to MS MARCO, a collection of datasets focused on deep learning in search that are derived from Bing queries and related data. Research organisations can gain access to the MS MARCO datasets instantaneously via the MS MARCO homepage.

	The MS MARCO dataset has been cited in over 1400 research papers since its release and has been used for a range of research issues, including in relation to misinformation and disinformation. Because the dataset is provided open source, the extent to which it has been used for disinformation related research purposes cannot easily be ascertained. However, the dataset has been cited in various academic papers concerning misinformation and disinformation, including: • "Retrieving Supporting Evidence for Generative Question Answering", SIGIR-AP '23: Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region, November 2023. • "Cross-Genre Retrieval for Information Integrity: A COVID-19 Case Study", In: Yang, X., *et al.* Advanced Data Mining and Applications. ADMA 2023. Lecture Notes in Computer Science), vol 14180. Springer, Cham, November 2023. • "Personas as a Way to Model Truthfulness in Language Models" New York University, ETH Zurich, et al. arXiv:2310.18168, October 2023.
MS MARCO Web Search	In 2024, Microsoft released the MS MARCO Web Search dataset, a large-scale information-rich Web dataset, featuring millions of real clicked query-document labels. This dataset closely mimics real-world web document and query distribution, provides rich information for various kinds of downstream tasks. MS MARCO Web Search further contains 10 million unique queries from 93 languages with millions of relevant labelled query-document pairs collected from the search log of the Microsoft Bing search engine to serve as the query set.
Other publicly shared datasets	Bing Search also offers use of Bing APIs to the public, which include services such as Bing Image Search, Bing News Search, Bing Web Search. Bing Search provides free access to these APIs for up to 1,000 transactions per month, which may be leveraged by the research community. Given the open nature of the Bing Search index and public nature of search results, researchers can use Bing to run specific queries and analyse results (unlike social media which may require private accounts or connections between users to access certain materials).

Elections & Societal Resilience team

Microsoft believes technology companies have a responsibility to help protect democratic processes and institutions globally. Though threats to democracy have always existed, the tactics of adversaries are constantly evolving. Microsoft is protecting open and secure democratic processes by providing services and technology to secure critical institutions, protect electoral processes from cyberattacks, and build public trust in voting procedures.

Microsoft's Elections & Societal Resilience team 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.

Beyond our commitment to combat deceptive use of AI during the electoral process discussed in this report, we implemented additional actions safeguarding candidates, election campaigns, election authorities, and voters:

- Microsoft's Campaign Success Team supported political parties and campaigns around the world to navigate the world of AI, combat the spread of cyber influence campaigns, and protect the authenticity of their own content and images.
- Microsoft's Election Communications Hub continued to support democratic governments around the world as they build secure and resilient election processes.
- Microsoft established a Virtual Situation Room, bringing together resources across the company to monitor, support, and protect elections in France and UK.
- Bing Search implemented a multifaceted approach to election integrity and integrated specialised answers and information panels for the elections across the European Union, with a link to official sources of information, which included voting information relevant to each EU Member State.

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 and on DIGI's website.

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

(O.7) Bing Search, MSN, Microsoft Advertising, LinkedIn

Microsoft On the Issues	Blog contains announcements on technology policy issues, including disinformation. For example, Meeting the moment: combating AI deepfakes in elections through today's new tech accord - Microsoft On the Issues, Securing US Elections from Nation-State Adversaries - Microsoft On the Issues and Expanding our Content Integrity tools to support global elections - Microsoft On the Issues.
Microsoft Reports Hub	Transparency reports include Digital Safety Content Report and Government Requests for Content Removal Report.
Microsoft Digital Defense Report 2025	Report encompasses learnings from security experts, practitioners, and defenders at Microsoft to empower people everywhere to defend against cyberthreats. Includes dedicated section on disinformation.
Microsoft Responsible AI Transparency Report 2025	Provides insight into how we build applications that use generative AI; make decisions and oversee the deployment of those applications; and learn, evolve, and grow as a responsible AI community.
LinkedIn Transparency Center	Community Report Government Requests Report
LinkedIn Blog	Blog contains information on actions to combat disinformation, including New LinkedIn profile features help verify identity, detect and remove fake accounts, boost authenticity, 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 fostering a trustworthy information ecosystem by enforcing our policies, driving research and innovation in emerging technologies, and collaborating with our partners, the academic community, and our users. This report outlines the measures being taken by Bing Search, MSN, Microsoft Advertising, and LinkedIn to mitigate and interrupt the spread of disinformation and misinformation. It also showcases the company's endeavours to fulfill the objectives and pledges of the Australian Code of Practice on Disinformation and Misinformation.