Microsoft submission on **Combatting Misinformation and Disinformation Bill 2023**

Submission to the consultation process run by the **Department of Infrastructure, Transport, Regional Development, Communications and the Arts**

20 August 2023

1 Introduction

Microsoft welcomes the opportunity to present this submission to the Department of Infrastructure, Transport, Regional Development, Communications and the Arts (the **Department**) in response to the exposure draft of the *Communications Legislation Amendment (Combatting Misinformation and Disinformation) Bill* 2023 (the **Bill**)[1].

At Microsoft, we are committed to advancing trust and security across our range of products and services, and recognise the importance of information integrity to the experience of our users. We want our users to feel genuinely empowered to make informed assessments about the information they encounter online, and recognise Microsoft has a crucial role in that process through efforts such as use of trust indicators and content provenance.

Combatting information operations that attempt to mislead or misinform our users is one of the many ways we seek to safeguard information integrity globally. We have demonstrated this commitment in Australia as a signatory of the voluntary *Australian Code of Practice on Disinformation and Misinformation* (**Voluntary Code**).

After considering the contents of the Bill, along with the accompanying **Guidance Note**[2], Microsoft has prepared this submission with the aim of supporting a version of the Bill that seeks to achieve its objectives in a more balanced, sustainable, and safeguarded manner.

[1] The Bill primarily amends the *Broadcasting Services Act 1992* (Cth) (**BSA**), with consequential amendments to several related pieces of legislation. A reference in this submission to a section of the Bill is a reference to that section as it appears in the proposed schedule 9 of the BSA, contained in 'Schedule 1 – Main amendments' of the Bill.

[2] Following the approach taken in the Guidance Note, we use the term 'misinformation' in this submission to refer to both misinformation and disinformation unless a distinction between the two is required, in which case we will make such distinction clear (or it will be clear from context).

In this submission, we address the following:

- Microsoft's approach to information integrity;
- Challenges in defining misinformation, operationalising those definitions, and specific definitional challenges in the Bill;
- The scope of impacted services, including the categories currently contemplated by the Bill and the need for additional excluded service categories;
- How the Bill handles transparency and the measuring of effectiveness, including specific information gathering concerns and challenges of measurability;
- Challenges raised by Misinformation Codes and Misinformation Standards; and
- Other observations and recommendations, including with respect to penalties, human rights and the need to incorporate greater oversight mechanisms into the Bill.

2 Microsoft's approach to information integrity

Microsoft recognises that creating a healthy information ecosystem is crucial to fostering a trusted and safe online environment. Through our global efforts aimed at understanding and responding to information operations, we have come to appreciate that there is neither a 'one size fits all' approach to this type of work, nor can it be viewed as the domain of technology providers alone. Fighting misinformation requires a whole-of-society approach, as well as governance responses that are proportionate and targeted in nature.

Microsoft launched the **Democracy Forward Initiative**[3] in 2018 to coordinate and track the efforts being undertaken across our organisation on protecting and strengthening democratic institutions. The Democracy Forward team leverages Microsoft's role as an enterprise software provider to increase our clients and users' ability to counter external efforts that compromise a healthy information ecosystem.

In June 2022, Microsoft announced our **Information Integrity Principles**. These principles were adopted across relevant Microsoft products and teams to ensure that we adopted an organisational approach to information integrity while still recognising the immense diversity of services, users, and risks across the company.

The following principles establish a foundational set of commitments that Microsoft teams can use to inform their policy, product development and risk assessment work:

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.*

[3] Previously known as the *Defending Democracy Program.*

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.*

Microsoft's commitments under the Voluntary Code

In Australia, we have produced annual disinformation and misinformation transparency reports pursuant to the Voluntary Code since its launch in February 2021. The Microsoft services covered by the scope of the Voluntary Code are:

- **Microsoft Advertising:** our proprietary online advertising network, which provides advertising displayed on Bing Search and 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, nor does it control the operation or design of the indexed websites or have an ability to control what indexed websites publish.
- **Microsoft Start:** a service which delivers licenced news and content across web and mobile for Microsoft customers and syndication partners.
- **LinkedIn:** a real identity online professional networking service for web and mobile, designed for professionals to connect and interact, grow their professional network and brand, and to seek career development opportunities. It includes user-generated content.

Trust indicators and informational action

A free and open internet is crucial to healthy, democratic societies. Unfortunately, the same fundamental structural conditions that enable the open flow of information critical to economic and societal growth can also enable the proliferation of low quality, and sometimes harmful, information online. As such, government, civil society, and industry collectively need to approach responses as an exercise in countering and managing the risks of misinformation.

At Microsoft, this dynamic informs our preference for informational responses that are both proactive in nature and respectful of the fundamental rights of our users. We have found that raising awareness for, and embedding signals of, trustworthiness and provenance is an effective way to foster a healthy information ecosystem online. Examples of this approach in practice include:

- **Search Coach** and **Search Progress** in Microsoft Teams enhance information literacy among students. Search Coach helps students form effective search queries and identify reliable resources in a safe, ad-free environment. It offers various filters and provides coaching tips to improve the information literacy of students. The app also includes a reliability indicator for each search result. Search Progress allows educators to assign a research project to their students, providing insight into a student's thinking process.
- **Democracy Forward Program**, an innovative effort to protect democratic institutions and processes from hacking, explore technological solutions to protect electoral processes, and defend against disinformation. This program leverages our role as a business and software provider to increase our clients' ability to counter outside efforts to compromise their security infrastructure.

- **Microsoft Threat Analysis Centre (MTAC),** established in July 2022 following our acquisition of Miburo Solutions, a cyber threat analysis and research company specialising in the detection of and response to foreign information influence operations. Working in close collaboration with the Microsoft Threat Intelligence Center (**MSTIC**), MTAC has enabled us to expand our threat detection and analysis capabilities to shed light on the ways in which foreign actors use information operations in conjunction with cyber-attacks to achieve their objectives.

3 Definitional challenges

Microsoft acknowledges that the Bill seeks to establish a more formalised regulatory framework that can operate over the long-term. To enable this, the Bill includes various definitions and scoping provisions that establish the definitional foundation for this scheme.

As these establish the legislative basis for potential future exercise of new powers by the ACMA, and more generally contribute to community and industry expectations around what misinformation is, Microsoft views these aspects of the Bill as the most important to get right from the outset.

We also recognise that the definitional challenges in the Bill are likely to be shared by similar co-regulatory or self-regulatory efforts at combatting misinformation online.

(a) Challenges in defining regulated speech

As with any public authority attempting to combat misinformation, the Australian Government is tasked with balancing its regulatory response with the fundamental rights of the community, including privacy, expectations of an open internet, and freedom of expression, of which freedom of political speech is a component. This challenge is most acute at a drafting level with respect to definitional provisions.

While we provide some commentary on specific drafting in the Bill at section (b) below, there are several overarching challenges that we wish to flag:

- **Developing subject area**: The nature of online misinformation, as well as understandings of how to best respond, are constantly evolving. Targeted research into misinformation, as well as associated fields of study (such as political science, psychology, sociology, information theory and media studies), continue to reflect divergent perspectives on the causes, sources, impacts of, and effective responses to, the challenge at hand.

  Co-regulatory and self-regulatory responses to misinformation can evolve in response to developing research and best practice over time, and this has been demonstrated in practice in recent years. Codifying such definitions into law has inherently longer-term implications and therefore definitional matters should be approached cautiously.

- **Subjectivity:** Fundamentally, efforts directed at reducing the dissemination of misinformation require some form of assessment of the truthfulness of content. While arbitrating the truth will, at times, be straightforward and noncontroversial, there are many settings where truth or falsity are harder to classify with confidence.

  Similarly, the terms 'misinformation' and 'disinformation' have a widely varied usage in the community; one person's conception of misinformation is another person's conception of political speech. The political forces at play in a given jurisdiction, as well as the enforcement

posture of independent regulators, can contribute heavily to perceptions of truthfulness and acceptable speech, which can also vary over time.

The Guidance Note makes clear that the Bill is not intended to give the ACMA any role in directly regulating individual pieces of content, nor determining what is considered truthful. We appreciate that such safeguards are designed to manage community expectations and the risks of government-regulated speech.

However, without the ACMA being able to at least contribute to dialogue on the truthfulness of content, disputes regarding whether content is misinformation or not (including disputes between service providers and the ACMA itself) will be difficult to resolve without judicial intervention. While involving impartial adjudicators like the judiciary may be productive in certain circumstances, it is an inappropriate default backstop given the scale of online information and the costs of and time involved in participating in formal legal action.

Given the inherent subjectivity that sits at the core of the proposed regulatory framework, the Department may wish to consider including in the Bill an appropriate mechanism for industry and the community to engage, in order to clarify instances of contested information integrity.

- **Context is key:** Under regulations targeting harmful or illegal speech, assessments carried out by service providers generally need to be performed on a case-by-case basis in order to ensure the context underlying a given piece of speech has been considered. Context is also crucial when assessing content that may present misinformation.

  Misinformation and (particularly) disinformation have been described as a "context-bound phenomenon",[4] with various contextual properties extrinsic and intrinsic to the content being potentially relevant to whether it meets a given threshold for regulated speech.

  Misinformation and disinformation require case-by-case assessments of content in the context in which they appear. When the nature of online content is factored in, this also means that an assessment performed on a piece of content in one context may not be transferable to the same content disseminated in a different context.

  Definitionally speaking, the Bill appears to accord with these principles. In addition to definitions of misinformation and disinformation, the Bill incorporates contextual analysis through the requirement for regulated content to reach a 'serious harm' threshold that factors in various considerations. Similarly, several of the Bill's exclusions rely on contextual analysis, such as exclusions for content produced in good faith for entertainment, parody or satire.

  Operationalising contextual analysis at scale is not only challenging for providers of digital platform services, but also risks undermining free expression online. This is largely due to the immense volume of online content and is also complicated by the fact that contextual inputs may not be apparent, available, or reasonably accessible to providers. For instance, consider the scenario where content flagged as misinformation requires consideration of academic material sitting behind a third-party paywall, or private information inaccessible to the provider, in order to confidently ascertain its truthfulness.

  At Microsoft, one of the ways we navigate the context-based phenomena of misinformation is to empower our users to better evaluate the reliability of the information they encounter

[4] Michael Hameleers, Disinformation as a context-bound phenomenon: toward a conceptual clarification integrating actors, intentions and techniques of creation and dissemination, *Communication Theory*, Volume 33, Issue 1, February 2023

online. We help our users assess information in context, such as by making content provenance more apparent, or by raising awareness for the markers of low-quality information. As another example, LinkedIn provides members with information to help them make more informed decisions about members with whom they interact through it's "About this profile" feature. This feature show users when a profile was created and last updated, along with whether the member has verified a phone number and/or work email associated with their account.

As the Department has also made clear that the Bill is intended to focus on systemic issues and processes, it will be important to consider how a systems-focused approach can be reconciled with the contextual nature of content analysis.

- **Coherence with international standards:** While there will be online misinformation concerns related to specifically Australian issues, a borderless internet makes the problem global in nature. In this context, there is value in achieving a level of international coherence where appropriate, particularly surrounding foundational definitions.

  In recent years, many jurisdictions around the world have proposed or adopted regulation aimed at combatting misinformation and disinformation. While this may be necessary to achieve some policy intents, implementing bespoke or excessively localised regulatory regimes in each jurisdiction is likely to lead to increased complexity for businesses of all scales and poor outcomes for end-users from the perspective of consistent information integrity.

  In many regulatory contexts, Microsoft sees international standards, particularly those established by recognised standards-setting bodies, as being important to achieving consistency and interoperability of online regulation. The ACMA has previously noted the importance of the Voluntary Code being informed by international developments in misinformation regulation.

  We recognise that, at present, no single definition of misinformation or disinformation has seen support at a global level. We also appreciate that there is an extent to which any agreed definition will still be complicated in practice by the inherent subjectivity of this issue.

(b) Specific commentary on definitions in the Bill

In addition to the above general commentary on definitions in the Bill, Microsoft makes the following more specific comments:

- Sections 7(1)(a) and 7(2)(a) outline that content will be misinformation and disinformation if it "**contains information** that is false, misleading or deceptive" (among other prerequisites). Using '*contains* information', as opposed to '*is* information' may unnecessarily complicate compliance efforts.

  As 'content' is intentionally defined very broadly and extends to include content in any combination of forms, it may be difficult for service providers to demarcate the line between different pieces of content. For example, a webpage can be considered content, but as can the various distinguishable pieces of content displayed on that webpage.

  If section 7(1) and (2) use 'contain' instead of 'is' as part of the threshold test, it may lead to unintended consequences in practice, such as service providers feeling obligated to remove content at the top level even though only a specific piece of sub-content is of concern.

Given the potential richness of information online, it is important that service providers are able to demarcate between different types of content in their compliance efforts.

- Sections 7(1)(d) and 7(2)(d) outline that content will be misinformation and disinformation if its provision on a digital service is reasonably likely to cause or **contribute to serious harm.** Unlike the Voluntary Code, which only references content that is reasonably likely to *cause* harm, we note that the Bill also extends to content that is reasonably likely to *contribute* to serious harm.

  The proximal relationship between content and serious harm is important to keep limited for practical reasons, as well as to ensure that providers are not incentivised to over-enforce which could come at the expense of coexisting user rights of privacy and free expression.

  While we appreciate that some content may have a less direct relationship to harm, contributing to serious harm alone provides too broad a test for providers to enforce around. For example, content may contribute to serious harm in a very minor way amongst other more material contributions. Further, it can be difficult to appropriately trace contributions by online content to real-world serious harm and may be mistakenly identified where content is merely related at a subject-matter level to a given example of serious harm.

  We recommend either removing 'contribute' as a relational factor, or inserting 'directly and substantially' before the words 'cause or contribute'.

4 Scope of impacted services

In our view, and given the other important rights that coexist in this context, the Bill should focus only on those services that are the most likely to facilitate the spread of online misinformation. Services that allow content to be shared publicly, that with a noted relationship to misinformation, or have technical functionality that contributes to the spread of misinformation, should be the exclusive focus of the Bill. For example, general purpose social services with heavily tailored content recommendation algorithms that enable the viral spread of content in a manner unlike other service types.

As presently drafted, the Bill would capture services where regulatory responses are generally impractical and inappropriate and where the risk to fundamental rights is greater, such as enterprise services. Also, the Bill's threshold of serious harm is highly unlikely to be met on low-virality services, such as closed or subject-specific communication services, or enterprise services. To the extent that misinformation does exist on such services, it will likely be of low and limited impact, with self-regulatory approaches by providers being the more proportionate means of responding if necessary.

(a) Scope of digital platform services

Appropriately defining and distinguishing between the different services that will be subject to the Bill is critical to reaching its stated aims.

While the Bill's definition of 'digital service' appears sound, the present scoping of the more material definition of '**digital platform service'** is likely to lead to confusion amongst regulated entities and appears misaligned with the aims of the Bill. Clause 7 of the Bill specifies that a digital service is a digital platform service if it is a content aggregation service, connective media service, media sharing service, or another service specified at a future time by Ministerial instrument.

While providing these subtypes of digital platform service is likely intended to support proportionate enforcement, our overall view is that the definitions are broad or inappropriately grouped. Microsoft generally advises against the grouping together of diverse sectors and considers a differentiated

approach to online regulation to be beneficial, both for industry clarity as well as for the underlying regulatory intent. Different sectors of the online industry, and the different services they offer, are used by different audiences, have varying levels of risk, and have different interventions available to respond to those risks.

Microsoft's specific feedback on each subtype under clause 7 of the Bill is as follows:

- **Content aggregation services:** Under the Bill, these are services that have a primary function of collating and presenting content to end-users from a range of online sources, with the Guidance Note outlining that this includes both search engines and news aggregation services. While we understand the role that each of search engines and news aggregation services play in a misinformation context, we query the appropriateness of grouping them together for the purposes of the Bill (and the regulatory instruments that may be introduced pursuant to the Bill in the future).

  While both of these service types connect users with information, from a utility and experience perspective, people generally use search engines and news aggregation services for quite different purposes and the services, in turn, are designed quite differently. For example, search engines respond to user queries and return relevant results, whereas news aggregation services play a greater role in curating content displayed, which may be partially personalised to account for geographic location, disclosed interests, and so on. Search engines source content through a process of collecting, indexing and ranking material available on the world wide web, whereas news aggregators are more likely to have a set of defined relationships with sources / syndicators of news content.

  Given that the various categories of digital platform service are likely to play a role in how the ACMA targets future exercises of its powers under the Bill, we query whether grouping search engines and news aggregators from the outset overstates their similarity and may lead to enforcement actions that are disproportionate to one or both service subtypes.

- **Connective media services:** These services have a primary function to allow end-users to interact with or link to each other online and include interactive features. This is a very broad category of services, which the Guidance Note acknowledges by stating that 'a majority of digital platform services will fall into this category'.

  While far from an exhaustive view of the services likely to be caught by the Bill's definition, the Guidance Note lists messaging services, social media, web-forums, dating sites, and online peer-to-peer marketplaces as examples of connective media services. In contrast, the same spread of services would fall under at least three separate service definitions under the *Online Safety Act 2021 (Cth)* (**OSA**), with different industry codes applying to each under that regulatory scheme.

  The risk factors ordinarily associated with the example services provided in the Guidance Note also diverge greatly. For instance, the spread of misinformation is far more pronounced on general-purpose social media where end-users often interact with 'the world at large' and content is able to spread widely due to popular use of the service in the community. On the other hand, interaction is much more targeted and information virality is lower on interest-based or professional networking platforms. For example, LinkedIn is a real-identity professional networking service, where each member has a LinkedIn profile associated with their LinkedIn account. Members must use their real name and include accurate information within their profile. Indeed, LinkedIn is part of its members' professional identity and has a specific purpose. Activity on the platform and content members share can be seen by current and future employers, colleagues, potential business partners and recruitment firms, among

others. Given this audience, members largely limit their activity to professional areas of interest and expect the content they see to be professional in nature.

Given that the subtypes of digital platform services should enable the ACMA to be in a better position to make differentiated responses to digital platform services, we query the usefulness of having such a broad category. It may be more appropriate to revise this definition to more closely align with definitions of social media.

The Department may wish to consider how similar definitions operate under the OSA. While we are not recommending that the Bill reflect a strict alignment with the OSA, we include the below commentary in order to illustrate how similar service distinctions have been handled in recent examples of digital law reform. The Bill's definition of connective media service and the OSA's definition of social media service, while similar, have the following critical differences:

Comparison	Commentary
Bill (s 7(3)(a)): 'a primary function of the digital service is…' **OSA (s 13(1)(a)(i)):** 'the sole or primary purpose of the service is…'	The Bill's drafting would allow for a service with several core functions to be in-scoped if one primary function was 'online interaction', even if that was not its predominant function or the overall purpose of the service. Function is also a much broader concept; services will usually have many more functions than they have purposes. The 'sole or primary' test in the OSA, and its focus on 'purpose' (rather than functions) of a service better targets its scope toward social media services and away from services that, while similar, have a distinct risk profile.
Bill (s 7(3)(a)): 'to enable online interaction between 2 or more end-users' **OSA (s 13(1)(a)(i)):** 'to enable online social interaction between 2 or more end-users'	Based on our reading of the Bill and associated Guidance Note, as well as our understanding of the broader misinformation landscape, we considered that 'connective media' should cover online *social* interaction as opposed to online interaction generally. Left unparticularised, online interaction captures essentially all interactions between more than one end-user, including those occurring in settings not ordinarily linked with misinformation. Unlike the OSA (at s 13(2)), the Bill does not provide any clarity on what 'online interaction' is intended to mean. The OSA also contains two statutory notes clarifying that 'online social interaction does not include (for example) online business interaction' and 'social purposes does not include (for example) business purposes'. Microsoft considers the distinction between general-purpose social media and other online interactive media, such as professional networking services, to be productive.
Bill (s 7(3)(c)): 'the service has an interactive feature' **OSA (s 13(1)(a)(iii)):** 'the service allows end-users to post material on the service'	The Bill's definition of interactive feature (s 5) is significantly broader than similar OSA concepts of posting material on a service. As with the above comments, the effect of this is that low-interaction services, or interactive services that are neither social nor general purpose in nature, will be in-scoped where a single interactive feature exists. For example, a movie rating website that makes the ratings of films by end-users observable by other end-users would satisfy the requirements of having an interactive feature (s 5(c)(ii)).

- **Media sharing services:** These services have a primary function to provide audio, audio-visual or moving visual content to end-users. Unless the service has an 'interactive feature' it will be excluded from the scope of the Bill.[5]

  The Guidance Note states that this service subtype may include podcasting services. However it would not include broadcast or streaming video-on-demand services. The Department has also communicated in industry briefings that this was not intended to capture file storage services.

  From the current drafting in the Bill, we believe that this subtype is likely to capture a wider variety of services than intended, including file storage services. While the existence of an interactive feature is meant to limit the scope impacted services, the definition of 'interactive feature' is too broad in nature to achieve this objective.

  For instance, a digital service will have an interactive feature where it 'allows end-users to post content on the digital service'. As 'post' or 'posted' includes an end-user causing content to be accessible to or delivered to one or more other end-user, with 'accessible' including access that is subject to a password. As a result, a service that, for example, enables end-users to privately store files online may be deemed a media sharing service as posting occurs and access by other end-users is possible, even though it requires a password or permission.

  In addition, as end-user is not defined to mean only individual end-users, a digital service that allows enterprise end-users to host an interactive portal on the service for internal use within the business may be captured (as an end-user has posted content (comprising the portal) and made this available to another end-user, being employees).

  The potential for this wide scope is particularly concerning given the unique human rights considerations at play in relation to personal storage services; content stored on such services is reasonably expected to remain private and is often intended for personal access (and not broad dissemination). In short, regulation of personal storage services significantly impacts individuals' realisation of their rights to privacy, freedom of expression, and freedom of opinion. Subject to certain overriding legal requirements, an individual's online file storage should be entitled to similar rights as their offline storage.

Improving the scoping of digital platform services will not only sharpen the regulatory intent of the Bill, but will also ensure that future exercise of ACMA powers are as proportionate and targeted as possible. Achieving this not only avoids regulatory uncertainty for providers, but also reduces the overall likelihood that the Bill, and compliance with its subsequent enforcement instruments, will undermine freedom of expression and privacy rights online.

We also acknowledge that the Bill allows the ACMA to exercise its information gathering, rule-making and code and standard powers in relation to a particular digital platform service or class of digital platform services. We recommend that future drafts could require the ACMA, when exercising a power

[5] However, we note that this does not appear to exclude such services from all of the ACMA's digital platform rule powers under Part 5 of the Bill's proposed Schedule 9 of the BSA.

over a class of services, to consider whether individually targeting particular digital platform services would be a more proportionate way to achieve the given objective.

(b) Excluded services

Microsoft acknowledges that the Bill already reflects several common-sense exclusions from in-scope services, particularly the exclusion of email services, SMS and MMS. However, when considering the nature of online misinformation, together with competing concerns around free expression and exchange of information online, we recommend that the Department consider expanding the exclusions to include enterprise services.

Taking a differentiated approach to services is critical to effective digital regulation, particularly when it comes to distinguishing the risks posed by consumer-facing services from those posed by enterprise services. Enterprise services are those that are provided to business or institutional customers (including government), for the purpose of supporting the business or operations of that customer, or for other 'B2B' purposes. There are extensive differences between enterprise and consumer services online which arise in various relational, technical, functional contexts:

Enterprise	Consumer	Impact on Bill
Small business, enterprise business, institutional and government customers	Consumer / retail customers	**As consumer services market toward individual consumer customers, they play a more direct role in safeguarding information integrity for Australian end-users. Enterprise services, on the other hand, primarily market toward business or other organisational customers, who may have their own internal information quality commitments, as well as the policies and procedures designed to meet those.** **Enterprise services may not even be available, accessible, or desirable to individual, consumer customers, reducing the need to impose increased regulatory obligations aimed at combatting community exposure to misinformation.**
Available to approved customers, lacks mass-market appeal	Available to world-at-large, low or no customer-approval	
Either not accessible to consumers or access barriers exist (resourcing, expertise, etc.)	Few or no barriers to consumer access	
Customer is not always end-user (for example, business procures enterprise service for use by its employees)	Customer is also the end-user.	**Providers of enterprise services are unlikely to have direct access to end-users, undermining their ability to comply with enforcement instruments actioned under the Bill, such as fielding end-user complaints, measuring misinformation identified on platform, etc.** **Oftentimes, providers of enterprise services will be unable to view discrete instances of content, let alone the surrounding context that is critical to consider when identifying misinformation.**
No direct relationship between provider and end-user (intermediated by customer)	Provider and end-user have direct relationship	
Contract with enterprise customer, but not end-users	Less formalised / negotiated agreements with end users	**Enterprise services are deliberately procured for specific purposes and providers are likely to have their own contracts with customers that may reflect an agreed position on information integrity and authentic behaviour.**
Consideration in the form of service fees or subscriptions	Free, low cost, freemium' or ad-funded services	
Small scale distribution potential	Wider scale distribution potential	**General-purpose consumer services can attract wide usage within the community and are therefore attractive platforms for disinformation campaigns to take root. Enterprise services are considerably less likely to harbor this potential.**

Overall, Microsoft views the risks posed by enterprise services in a misinformation or disinformation context to be very different to the risk potential of certain consumer services. Further, there are considerable distinctions with regard to customer and end-user relationship, which directly impact the ability for providers of enterprise services to comply with regulated speech requirements.

5 Transparency and measuring effectiveness

Microsoft recognises the role that transparency plays in monitoring the state-of-play when it comes to online misinformation and related responses from the tech industry. For several years now, Microsoft has published transparency reports in relation to this space, including both voluntary and mandatory measures relating to misinformation, as well as voluntary disclosure of responses to major international information operations.

(a) Information gathering and record-keeping rules

The Bill provides the ACMA with new information gathering powers, including the ability to set record-keeping rules that apply to digital platform services, or issue notices requiring the production of documents or information. Together, these provisions would enable the ACMA to obtain information that it may consider has not been adequately disclosed to date by signatories to the Voluntary Code. While Microsoft appreciates this desire, we also make the following observations and recommendations:

- **Consistency:** Commentary from the ACMA and other stakeholders in relation to the Voluntary Code has noted how the contents of transparency reports can differ widely between different digital platform services. In turn, it is suggested that this creates challenges for the ACMA and other persons wishing to understand and evaluate the state of misinformation in Australia.

  In our view, while better coordination amongst industry may be appropriate, transparency reporting can only be truly consistent to the extent the underlying services themselves are consistent in nature. Unlike other regulated sectors such as utilities, telecommunications and broadcasting which, at a general level, provide analogous services to customers, the 'online industry' is heavily segmented and diverse in terms of functionality, userbase, purpose and risk.

  This comes back to taking a differentiated approach to the online industry, acknowledging the different categories of services (i.e. social media being functionally distinct from search engines), as well as the different types of services within a given category (i.e. general-purpose social media being functionally and purposively distinct from professional networking).

  We recommend that the Bill acknowledge this practical reality, such as in the drafting of section 14(2) of the Bill. This could require the ACMA, before making a digital platform rule, to consider the extent to which specified digital platform services are consistent and whether the rules to make or retain records will be able to apply to the specified digital platform services in a proportionate and fair manner.

  Ultimately, we are also concerned that reporting requirements that are unduly or artificially consistent will lead to low quality data outputs that are less meaningful to understand and respond to misinformation online.

- **Existing reporting:** We also note that the ACMA's rule-making powers are paired with related reporting powers, which can be used to require digital platform services to report certain information to the ACMA in a form and manner specified. Noting the extensive transparency reporting done by some digital platform services (including disclosures made on an entirely independent basis), the Bill ought to require the ACMA to consider the extent to which information being sought under a reporting-related digital platform rule is already made available by any of the digital platform services within the scope of that rule. Not only will this reduce regulatory overhead and overlap for providers, but will also continue to incentivise voluntary reporting.

- **Disclosure by the ACMA:** The Bill allows the ACMA to publish on its website any information that it has obtained under Part 2 of the Bill. Given the breadth of information that could be obtained under such powers, we query whether such publication rights are appropriate or necessary in order to meet the policy objectives of the Bill.

  We acknowledge that the Bill requires that any 'digital platform rules' that compel reporting by providers to the ACMA must allow for providers to identify information in the report the publication of which could be expected to prejudice materially the commercial interests of a person. However, such a right is not incorporated into the general information gathering powers provided to the ACMA under the Bill.

  Where the ACMA intends to publicly disclose on its website information that it has obtained pursuant to a power under Part 2 of the Bill, providers are to be given advance notice of proposed publications and an opportunity to make 'submissions' in relation to the proposal, including identifying any commercially prejudicial information, before a date specified by the ACMA.

  In our view, current drafting on this front does not adequately address the issues associated with the public disclosure of information obtained via the exercise of regulator powers and renders the drafting out of step with other comparable statutory information gathering powers. While the ACMA is required to consider provider submissions under section 26 of the Bill, there are insufficient guardrails in place to ensure that the ACMA proportionately takes such submissions into account prior to publication.

  In addition to protections for commercial sensitive information, we also recommend that the Bill expressly allow providers to request non-disclosure of reported information for public safety and platform integrity grounds. In the context of online safety and preserving democratic values online, some platform interventions are kept confidential to ensure that they are not manipulated, circumvented or gamed by bad actors. This is particularly critical in the context of disinformation and other sophisticated information operations. The public disclosure of certain information could lead to the diminished effectiveness of efforts aimed at reducing misinformation.

  We recommend that the Bill:

  - require that, prior to publishing information asserted to prejudice commercial interests, the Authority of the ACMA be formally tasked with weighing up a provider's submission against the relative public interest in disclosure of that information;

  - make clear that providers have an ability to contest the publication as proposed (instead of merely making submissions disconnected from a clear objective);

  - expand the matters that providers are expressly able to identify in submissions to include confidential information (i.e. not restricted to commercially prejudicial information), information that is materially sensitive due to public safety and platform integrity concerns, and government sensitive and classified information;

  - further specify the nature of the publications that the ACMA can make under section 25 of the Bill, such as by limiting this to summaries of information reported or obtained under Part 2 of the Bill (instead of the ability to provide information or documents in full); and

- set a reasonable minimum timeframe for providers to respond to a proposed public disclosure of information by the ACMA under section 25 and 26. This period of time should be considered by the ACMA as reasonable in the circumstances, but in any case, be a period that is no less than 30 days.

(c) Measuring effectiveness

Transparency is also inherently linked to the aspects of the Bill that require measuring effectiveness. The Guidance Note makes clear that the Bill is intended to establish a graduated enforcement scheme for the ACMA that can be deployed over time in response to its assessment of how effectively a given digital platform service, or class of digital platform services, is dealing with misinformation and disinformation. On this topic, we make the following observations:

- **Defining effectiveness:** In our view, it is inherently difficult to meaningfully assess the effectiveness of measures aimed at responding to online misinformation.

  With other forms of regulated online content, such as content that is specifically illegal to possess or distribute, providers are better able to establish clear user-facing policies and internal governance procedures that adopt a zero-tolerance approach to violative content. For some such content, no level of prevalence would be deemed acceptable to the provider, and the effectiveness of its responses could be gauged by measuring the rate of proactive removal versus responsive removal (i.e. removal in reaction to a valid complaint).

  Responding to online misinformation is less clear cut, and in turn, cannot be measured in the same way as other regulated content. It is an area where providers should be able to apply the measures that are most contextually appropriate in light of the nature of their services. Indeed, as is acknowledged in the Bill's use of a 'serious harm' threshold, interventions into lawful free expression online must be reserved only for content reasonably likely to cause or contribute to serious harm. As a result, providers carefully weigh trade-offs between lawful freedom of expression and the likelihood of serious harm in monitoring and enforcing misinformation policies. For example, content removal is one of several response options, and one which many providers reasonably keep reserved for instances of harmful disinformation or misinformation. Other options include limiting the distribution of the content, for example in instances where the content is not likely to cause a serious harm but may still be desirable to limit in order to safeguard the interests of other users on the platform. Mitigating harm needs to always be balanced with maintaining an open internet that respects the fundamental rights of users.

  Informational responses, such as interstitial notices, trust indicators, community corroboration and media literacy campaigns, may also play a meaningful role in responding to misinformation. These measures contemplate coexisting with potentially false, misleading and deceptive information, rather than eliminating this type of content outright. As a result, their effectiveness cannot always be measured in quantifiable (or consistent) ways.

- **Key performance indicators:** Similarly, we understand that the ACMA has previously expressed a desire for signatories to the Voluntary Code to establish a framework of key performance indicators (**KPIs**) aimed to measuring the effectiveness of responses to misinformation. The Guidance Note reiterates this, noting that the Bill would enable comparison of metric and KPIs across the industry, leading to 'increased comparability'.

  We caution against drafting that would allow the ACMA to create uniform KPIs for digital platform services. As previously noted, digital platform services vary significantly in terms of functionality, purpose, userbase, content distribution model and exposure to misinformation

risks. Accordingly, responses aimed at reducing misinformation or increasing information integrity for users also varies significantly.

Requiring disparate services (even those who may share some characteristics) to conform to uniform or inflexible KPIs is unlikely to provide the ACMA with meaningful insight and may unintentionally incentivise compliance responses from platforms that ignore the nuances and complexities of their individual services. For example, time-based KPIs that measure how long a provider takes to remove a piece of content would be unworkable in this context and would incentivise automated, low-context enforcement.

Where KPIs are mandated by the ACMA, we recommend that these still be flexible, principles-based and connected to measures of effectiveness that are appropriate in the context. For example, a requirement that services measure appropriate qualitative signals relating to user perceptions of low-quality information on a service.

- **Measurability:** Another consideration to keep in mind is the extent to which certain aspects of misinformation, as a complex social and technological phenomenon, cannot be measured well or at all.

  The Bill could recognise this by permitting, or in some instances requiring, the ACMA to take into account findings from more participatory regulatory fora, such as industry roundtables and best-practice cooperation initiatives, including with representatives from civil society and other community stakeholders.

6 Misinformation Codes and Misinformation Standards

We understand that the Bill is intended to set out long-term framework that provides the ACMA with optionality as to its enforcement activity. The Department has stated that the new ACMA powers are intended to be exercised in a graduated manner in response to misinformation challenges that may arise over time.

Despite this intention, we consider that the Bill fails to establish sufficient guidance or guardrails for the ACMA's exercise of its proposed new powers, and as such, creates unreasonable uncertainty for industry. Below we outline several challenges we have identified with how the Bill presently handles the code and standard making powers of the ACMA.

- **Trigger points:** At present, in order for the ACMA take certain actions regarding codes and standards it must be satisfied that it is **necessary or convenient** to do so, in relation to codes, 'to prevent or respond to misinformation or disinformation on digital platform services' or in relation to standards, 'in order to provide adequate protection for the community from misinformation and disinformation on the services'.

  The biggest issue we see with this current drafting is that the ACMA may exercise powers as '**convenient**', even if it could not be argued that such actions were necessary. In our view the exercise of these powers should only be where necessary and proportionate to specific risks, consistent with rule-of-law principles.

  While we appreciate the need for regulators to have a reasonable degree of flexibility and discretion as to the exercise of their powers, we note that the regulation of what is primarily lawful speech ought to necessitate stricter guardrails for enforcement escalations, particularly in light of the government's duty to protect freedom of expression.

- **Selective application:** The ACMA has previously noted that more formal regulatory options would be particularly useful for "platforms that choose not to participate in the [voluntary] code or reject the emerging consensus on the need to address disinformation and misinformation."

  We support the targeted deployment of code and standard making powers toward contexts and providers most acutely associated with serious misinformation risks. Not only will this incentivise the participation in, and continuous improvement of, voluntary responses from industry, but will also ensure that government intervention into matters of lawful speech are kept at necessary and proportionate levels.

- **Roles of industry associations and the ACMA:** The respective roles of industry associations and the ACMA when it comes to code development should be appropriately outlined in the Bill. Due to the highly differentiated nature of the online industry, unlike other sectors there is no single industry association which represents all relevant organisations. Presently, the Bill contemplates industry associations providing industry participants with an opportunity to review and make submissions with regard to a draft code, but does not require the involvement of industry participants in the development itself. Given the uniquely diverse nature of the online industry, industry participants need to be offered more than time-limited consultation rights, but a genuine ability to help shape development of industry codes.

  Given that the Bill contemplates industry association-led codes, and regulator-led standards, and enables the ACMA to guide the scope of requested codes, it may be appropriate to establish guardrails in relation to the further ongoing involvement of the ACMA in developing the content of industry codes. At minimum this could include defining a set number of feedback exchanges between industry and the ACMA may avoid protracted and resource-intensive development processes, as well as a public consultation period of at least 42 days that are carried out prior to the wider public consultation required by the Bill to enable meaningful engagement from stakeholder groups.

  For a more robust approach to code making, the Department should consider a well-governed forum open to industry, academia, and government experts in which all parties have an equal voice. Standards Australia is well positioned to establish, govern, and manage such a forum to ensure it is open, transparent and allows all voices to be heard. If aligned with the principles of the World Trade Organization in establishing standards, then such a code could also be used to drive a more global adoption, enabling Australia to lead the way a code that could be adopted by both Australian and international service providers, reducing maintenance and implementation costs of the code.

7 Other observations and recommendations

While the previous sections of our submission cover the main issues Microsoft wished to raise, we also make the following final observations in relation to the Bill:

- **Enforcement and penalties:** We appreciate that any regulatory scheme must be accompanied by an appropriate incentive and deterrence structure. However, we query whether the enforcement mechanisms and related penalties set out in the Bill are proportionate or appropriate.

  Challenges raised elsewhere in this submission, such as the subjectivity of harm assessments or the difficulty in agreeing on what effectiveness looks like, ought to encourage a rethink on how enforcement and penalties are embedded into the Bill. For example, the Bill presently

exposes industry participants to civil penalties, at the discretion of the ACMA, for any non-compliance with a code or standard. Given the potential scope of either regulatory instrument in the future, such a blunt approach to penalties may not appropriately reconcile with codes and standards encouraging, for example, best effort actions, transparency reporting, or investments into longer-term, systemic improvements.

It may also have unintended consequences for those tasked with developing industry codes, who may be wary to develop robust requirements if they are to be exposed to unsuitably rigid statutory enforcement options.

- **Balancing human rights online:** While combatting misinformation and disinformation can further the pursuit of universal enjoyment of fundamental rights, such as by combatting narratives that would threaten the safety of minority or vulnerable populations or disrupt the right to health, these efforts can also threaten the rights to freedom of expression and privacy.

  Legislation and regulations addressing misinformation and disinformation must be necessary and proportionate, and carefully considered for risks of potential rights infringement, including where—as here—regulations may be regarded as government limiting individuals' freedom of expression (including the right to access information) or privacy through the actions of intermediaries (i.e. digital services platforms). In the above sections, we identify aspects of the Bill that may raise human rights considerations, including the "serious harm" standard, the challenges of determining the truth or falsity of content (and therefore whether it constitutes misinformation or disinformation), and the inclusion of personal storage services in the scope of the Bill.

  There are significant tensions between moderating content and protecting fundamental rights. We encourage mindful attention to competing rights, laws, and goals as the Bill progresses, as other states, regulators, industry, and civil society stakeholders may look to Australia's example of leadership on misinformation, disinformation, and human rights.

- **Incorporating greater review and oversight:** Finally, and considering the issues addressed in this section in particular, we query whether the Bill ought to incorporate stronger review and oversight mechanisms.

  Other examples of digital law reform have incorporated statutory review requirements on set timeframes. This not only inserts productive independent and Parliamentary review into the overall regulatory regime, but also accounts for the fact that the Bill is addressing a challenge where research, consensus and best practice are in a state of ongoing flux.

8 Conclusion

Safeguarding the integrity of information online is a shared objective of industry, government, and civil society. While technology plays a role in making content available to users, misinformation is a fundamentally human, global, and deeply context-based phenomenon that in turn requires a whole-of-society response.

Together with appropriate regulation, a strong and resourced civil society, public investment in information literacy, increased consumer and citizen awareness, and ongoing academic research all play a role in building a healthy information ecosystem.

Microsoft thanks the Department for the opportunity to contribute to this important consultation. We hope that the observations in this submission can assist in crafting regulatory responses to misinformation that are proportionate and strike the right balance with the rights and interests of the community.