Microsoft submission on **Safe and Responsible AI in Australia**

Submission to the consultation process run by the **Department of Industry, Science and Resources**

4 August 2023

Summary of views

Australia is especially well-positioned to benefit from the extraordinary social and economic opportunities created by the latest wave of AI technologies. To realise these opportunities, however, effective guardrails are required to promote the trustworthy and responsible development and use of AI systems. To that end, Microsoft supports the advancement of AI governance in Australia through:

1. Creating a regulatory architecture that reflects the AI technology architecture by tailoring the right regulations for each level of the tech stack.

2. Improving domestic coordination on AI regulation and policy, including both the layers of existing economy-wide and sector-specific regulations as well as overlapping reform processes in areas including privacy, cyber security, and online safety.

3. Advancing a risk-based approach by placing guardrails on AI systems that have the potential to meaningfully impact the public's rights, opportunities, or access to critical resources or services as well as to require effective safety brakes for AI systems that control critical infrastructure.

4. Continuing to collaborate in broader international efforts to work towards internationally coherent AI governance, such as continuing to participate actively in developing international standards; leveraging the NIST's AI Risk Management Framework; and contributing to the development of a voluntary international AI Code of Conduct as discussed at the G7 Summit in Hiroshima.

5. Promoting opportunities to benefit from AI technologies by raising community awareness and trust in AI through a transparency-led approach and addressing known regulatory blockers.

Microsoft appreciates the opportunity to comment on the '*Safe and responsible AI in Australia*' discussion paper (**Discussion Paper**) from the Department of Industry, Science and Resources (**DISR**).[1] As a provider of cloud services and other technology solutions in Australia, and as company operating at the forefront of innovation when it comes to artificial intelligence (**AI**), Microsoft is committed to

---

[1] Department of Industry, Science and Resources, Safe and responsible AI in Australia Discussion Paper, 1 June 2023.

working alongside Government to 'meet the moment' when it comes to safe and responsible AI.

Microsoft has welcomed past opportunities to provide our views to Government consultations relating to AI, including in the development of Australia's AI Action Plan; as a participant in the pilot of Australia's AI Ethics Principles; to the Data61 discussion paper on AI; to the Standards Australia AI Road Map; and in the Australian Human Rights Commission's Technology and Human Rights Project in 2019 and 2020. In 2018, Microsoft commissioned Future Eye to develop work around Creating a Social License for AI to Flourish, as well as a subsequent report in 2019 in which DISR participated.

We believe our experience as a developer, operator, and user of AI technologies provides a useful perspective to contribute to the development of AI policy in Australia. We look forward to continuing our engagement with the Government on this important issue.

1 Australia's AI Opportunity

Since the invention of the printing press, technological inventions have been driving exponential economic growth at an accelerating rate (see **Figure 1** below). Like the inventions that have come before it, AI has huge potential to help advance thinking and learning to improve the human condition. And in many ways, Australia is especially well-positioned to realise the extraordinary social and economic opportunities created by a wave of new AI technology due to our well-developed IT infrastructure, high adoption of cloud services, and strong commitment to free trade and international cooperation.[2]

Figure 1: Technology and GDP growth[3]
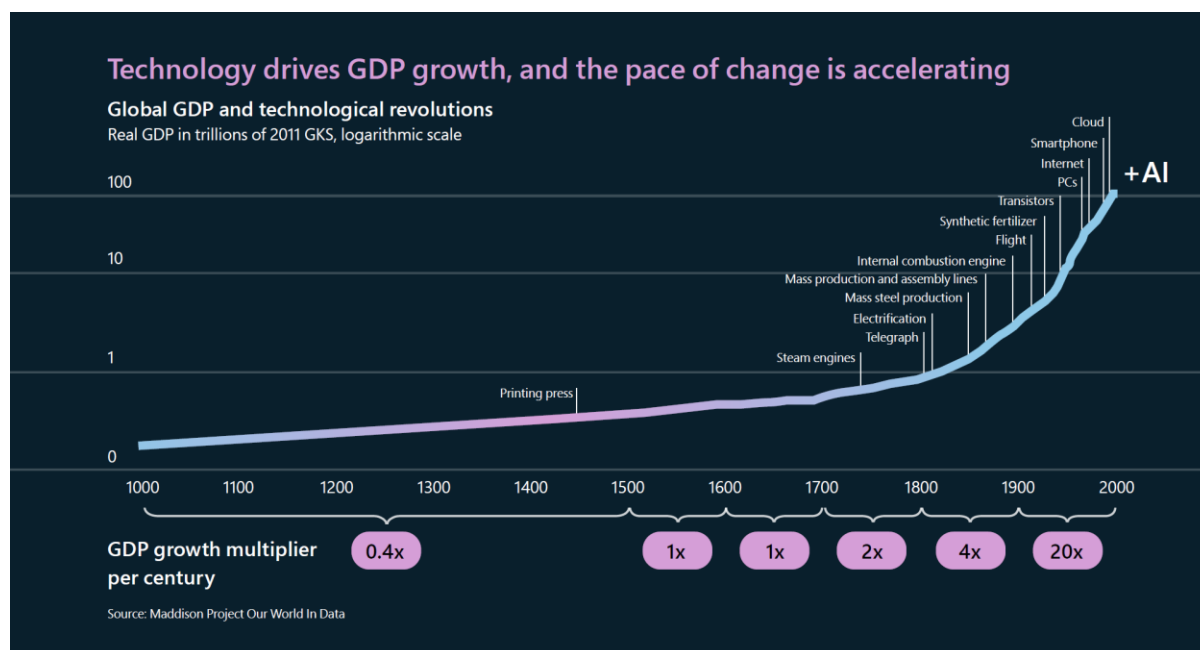
Australia's highly-skilled workforce is also poised to gain a significant productivity boost from a wave of new AI technology. Recent research by Microsoft and the Tech Council of Australia estimates that up to 44% of Australian workers' task hours could be automated or augmented by generative AI, and

---

[2] Business Software Alliance, BSA Global Cloud Computing Scorecard – Australia Country Report, 2018.
[3] Microsoft, Governing AI: A Blueprint for the Future, May 2023.

with the right adoption rates, could contribute $115 billion a year to Australia's economy by 2030.[4] 70% of this value would come from productivity gains.

This research also found that Generative AI is already copiloting work in a diverse range of job tasks including software development, marketing and sales, research, and management.[5] In software development, for example, the rapid adoption of generative AI tools has already significantly boosted productivity. A recent global survey found that 70% of developers are using or plan to begin using AI tools in their software development.[6] Studies have also found, for example, that developers using GitHub Copilot are between 30% and 55% more productive at writing software.[7] These tools are boosting individual wellbeing too – a survey of GitHub Copilot users found that 60% reported increased job satisfaction and 74% reported being able to focus on more satisfying work while using the AI tool.[8] Increased software productivity is expected to decrease the cost of software and thus realise more effective demand, increasing the pace of digital transformation.

In addition to accelerating productivity growth, the development and adoption of AI technologies has huge potential to support Australia's future economic diversification to build a resilient nation. Research by the Tech Council of Australia has found that digital technologies, including AI, are an increasingly essential part of the value chain of diverse businesses across the broader economy.[9] As recently recommended by the Productivity Commission, 'Innovation policy should broaden and give more emphasis to the spread and adoption of new technology and best practice. Adoption of digital technology, such as AI, and the better use of data by businesses can boost productivity and should be encouraged by government action.'[10]

2   Microsoft's commitment to safe and responsible AI

Since late last year, developments in AI technologies have captured the human imagination and created new tools to advance human learning and thought. At Microsoft, we have been excited to witness and contribute to this latest wave of technological progress. In many ways, we have been preparing for this moment for the better half of the last decade. Today, Microsoft over 350 people working on responsible AI across the organisation, implementing best practices for building, and helping our customers build, AI systems that are safe, secure, transparent and designed to benefit society.

The resulting advances in our approach have given us the capability and confidence to see ever-expanding ways for AI to improve people's lives. We've seen AI help save individuals' eyesight, make progress on new cures for cancer, generate new insights about proteins, fending off cyberattacks and helping to protect fundamental human rights. We've also seen AI's ability to act as a copilot assist with everyday activities, such as turning search into a more powerful tool for research and improving productivity for people at work.

---

[4] Microsoft and Tech Council of Australia, Australia's Generative AI Opportunity, July 2023, p3.

[5] Microsoft and Tech Council of Australia, Australia's Generative AI Opportunity, July 2023, p3.

[6] See Stack Overflow, 2023 Developer Survey.

[7] 30% is from an observational study of nearly 1 million developers writing code: see Thomas Dohmke, Marco Iansiti and Greg Richards, Sea Change in Software Development: Economic and Productivity Analysis of the AI-Powered Developer Lifecycle, 26 June 2023. 55% is from a controlled experiment on a task to implement a functioning software system: see Sida Peng et al, The Impact of AI on Developer Productivity: Evidence from GitHub Copilot, 13 February 2023.

[8] Eirini Kalliamvakou, GitHub Blog, Research: quantifying GitHub Copilot's impact on developer productivity and happiness, 7 September 2022.

[9] Tech Council of Australia, The economic contribution of Australia's tech sector, August 202, p6.

[10] Productivity Commission, Advancing Prosperity, Recommendations and reform directives, 7 February 2023, p17.

Since 2016, our approach to responsible AI has been grounded in **six foundational principles** of fairness, reliability and safety, privacy and security, inclusiveness, transparency, and accountability (See **Figure 2** below). We continue to build out our responsible AI program to put these principles into practice, with input from multi-disciplinary teams across Microsoft from research to engineering to policy, with oversight from executive leadership and the Microsoft board. This is a global investment in our responsible AI program with dedicated resources based in and supporting Australian customers and businesses..[11] This includes formalising these principles in our own internal Responsible AI Standard, the latest iteration of which we published in June 2022. The Responsible AI Standard acts as a framework to guide how we build AI systems. It also demonstrates the commitment we have at Microsoft to share learnings, encourage transparency and support our customers and the broader community to emulate similar models themselves.

These foundational principles are also aligned with DISR's AI Ethics Principles. This demonstrates our shared aim to implement appropriate safeguards to promote the safe and responsible use of AI to have a positive impact on society.

Figure 2: Microsoft's responsible AI principles
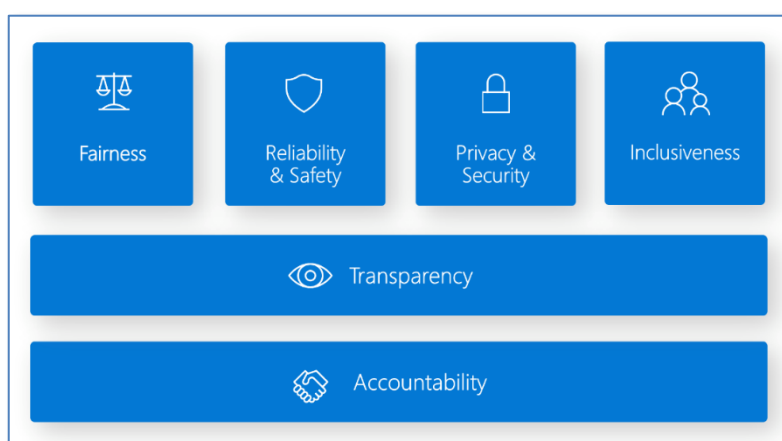
3   Governing AI in Australia

The enormous opportunities and new challenges of AI have attracted the attention of governments worldwide, prompting important questions about the adequacy of existing governance models to capture the opportunities while managing the challenges. We recognise that there is a need to manage the potential risks associated with the use of AI and to support its responsible adoption to ensure Australia benefits from the extraordinary opportunities of this technology. This submission sets out Microsoft's proposals to achieve both of these aims.

(a)   The importance of targeted definitions aligned to global best practice

A standardised set of core concepts and definitions is essential to ensuring a coherent regulatory framework across the varied sectors of the economy. It would not make sense, for example, to develop divergent definitions of AI or foundation models that apply only to Australia. We therefore recommend that the Government consider aligning core definitions with international best practice, which would help shape a globally consistent AI regulatory regime so that Australian organisations of

---

[11] See further Microsoft AI.

all sizes can continue to collaborate with others across borders in developing and using world leading technology.

For example, the Organisation for Economic Cooperation and Development (OECD) and National Institute for Standards and Technology's (NIST) AI Risk Management Framework definitions of an "AI system" is a strong definition that is being included in a number of proposals across different jurisdictions, including in the European Parliament's position on the EU's draft AI Act.

The definitions for foundation models should focus in on the way in which these models are a narrow class of AI models, capable of performing a range of tasks, intended to be adapted and integrated into a wide variety of downstream applications rather than being deployed directly to an end user. Definitions for developer and deployer should focus on those entities that control the design and development of an AI system and the entity that takes a decision about how and where to use an AI system, respectively.

(b)   Regulating at the right level of the AI technology stack

We believe there will need to be a legal and regulatory architecture for AI that reflects the technology architecture for AI itself. That is, AI governance should consider the different types of responsibilities that exist at three key layers of the technology stack: the **applications layer**, the **model layer**, and the **infrastructure layer**.

Although there is no single correct way to describe an AI tech stack, we consider that the diagram at **Figure 3** below is a good starting point. The third layer in the middle of this stack represents advanced pretrained AI models such as GPT-4 and is created by OpenAI based on the two layers below it: machine learning acceleration software and, in the case of GPT-4, AI datacentre infrastructure that Microsoft built in Iowa, USA.

Figure 3: Technology and GDP growth[12]
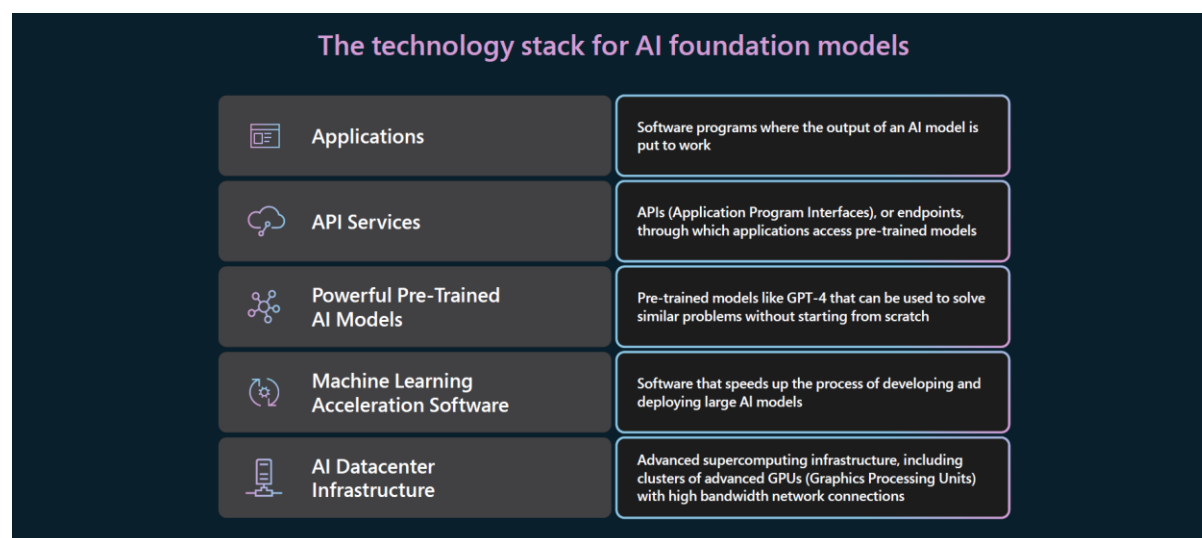
The **applications layer** is the top of the stack where businesses and consumers are deploying and engaging with AI systems. This is the layer where the safety and rights of people will most be impacted, especially because the impact of AI can vary markedly in different technology scenarios. As discussed further below, there are numerous existing laws and regulatory mechanisms that apply at the application layer.

In considering regulatory interventions in the applications layer, it is important to first consider the extent to which existing legal protections apply. In many areas, we may only need to apply and enforce existing laws and regulations, helping agencies and courts develop the expertise needed to adapt to new AI scenarios. Any risks of harm that emerge at this applications-level should be addressed in a risk-based way by supplemental reforms if there appears to be a gap not filled by existing regulation.

At the **pre-trained AI model layer**, we have seen a new class of highly capable foundation models emerge that are trained on internet-scale datasets and are effective out-of-the-box at new tasks. The capabilities of these foundation models are at once very impressive and can be harder to predict. Substantial thought, discussion, and work will be needed to define the appropriate threshold for what constitutes a highly capable AI model. Microsoft, along with other leading AI developers, will continue to share our specialised knowledge about advanced AI models to help governments define this regulatory threshold.

Should the Government consider that developing or deploying a highly-capable AI model should require licensing requirements, Microsoft will actively participate and contribute to efforts to define the requirements that must be met in order to obtain a license. An effective licensing regime should be designed to achieve safety and security objectives, with requirements such as advance notification of large training runs, comprehensive risk assessments, extensive pre-release testing by internal and external experts, multiple checkpoints along the way, and ongoing monitoring post-release. There should be a framework for close coordination and information flows between licensees and their regulator as well as international interoperability with regimes in countries with shared safety and security goals.

Finally, the **infrastructure layer** underpins the training and ongoing deployment of AI models and applications. Noting that datacentres in Australia are currently regulated by a network of security and safety controls, Microsoft will also support the Government in considering additional licensing requirements on the operators of AI datacentres used for the testing or deployment of highly capable foundation models. It may also be appropriate to consider targeted regulatory intervention in the context of autonomous systems that manage critical infrastructure.

This layered approach to AI governance builds partly on the 'Know Your Customer' principle used in financial services regulation that requires financial institutions to verify customer identities, establish risk profiles, and monitor transactions to help detect suspicious activity. Applied to AI services, it would make sense to apply a KY3C approach:

- First, developers of powerful AI models should '**know the cloud**' on which their models are deployed

- Second, in some scenarios (e.g. sensitive uses), the cloud services provider should '**know the customers**' who are accessing the powerful AI models

- Third, the public should be able to '**know the content**' that is created by these powerful AI models with labelling requirements for content to be labelled in important scenarios.

(c)   Best practice approaches to regulating AI

Microsoft believes that basing AI governance approaches around outcomes and risk is key to crafting and deploying regulatory models that are fit-for-purpose and sustainable in practice.

1. **Outcomes-based approach:** Through our experience with AI systems and AI regulation around the world, we have learned that focusing on outcomes and processes can be more effective in ensuring responsible AI use than prescriptive measures. This is particularly relevant given the huge potential for AI to be used in many different ways by many different organisations as well as its rapidly evolving nature.

   Moreover, an outcomes-based approach would encourage the adoption of evolving best practices and state-of-the-art tooling, both of which are crucial for effective compliance over the long-term. The right mix of outcomes-focused regulation and appropriate monitoring has the potential to foster a culture of continuous improvement within the AI ecosystem.

   Of course, focusing on outcomes requires a level of consensus on those intended outcomes. An outcomes-based approach will be most effective where it reflects shared values, objectives and concerns. One way of supporting a coherent outcomes-based approach would be to establish an international, voluntary code of conduct (as discussed further in Section 3(ii)).

2. **Risk-based and proportionate approach:** We support a risk-based approach to develop AI governance that is proportionate, differentiated and responsive to the risks of a given AI system. Risk-based regulation helps to appropriately target interventions toward high-risk applications of AI while still enabling responsible innovations and encouraging the beneficial uses of AI in the community.

   Risk-based approaches to AI regulation have received widespread support from regulators and stakeholders around the world. For example, the EU's draft AI Act sets out a risk-based approach with defined processes around risk identification and mitigation. Similarly, the NIST's AI Risk Management Framework was recently developed to help organisations manage AI risks and promote the trustworthy and responsible development and use of AI systems.

   It is crucial to consider the specific context and use-case when assessing risk in AI systems. For example, there will be many AI systems used even within critical infrastructure sectors that are low risk and do not require the same depth of safety controls such as employee productivity tools and customer service agents.

Facial Recognition Technology Model Law

Facial recognition technology raises issues that go to the heart of fundamental human rights protections like privacy and freedom of expression. These issues heighten responsibility for tech companies that create these products and this is why we believe they also call for thoughtful government regulation and for the development of norms around acceptable uses.

In Australia, the University of Technology Sydney Human Technology Institute (**HTI**) has proposed the widely commended Facial Recognition Technology (FRT) 'Model Law' which is risk-based and grounded in international human rights law. The FRT Model Law acknowledges the importance of context and the potential for responsible FRT use, and establishes a clear framework to support developers and deployers to do so.

(d)   Prioritising domestic policy coordination

As the Discussion Paper Notes, there are existing efforts from across the private and public sector to ensure responsible use of AI, appropriate procurement of AI systems, and encourage investment. These include an extensive range of **economy-wide laws**, including in relation to privacy, consumer protection, competition, corporations, and anti-discrimination. There are also numerous examples of **sector-specific regulations**, such as those governing telecommunications, broadcasting services, airline safety, motor vehicles, financial services, legal services, and medical devices.

We further note that there are **several overlapping reform processes already underway** that will impose rules on use of AI in Australia such as the Privacy Act Review, Australia's Cyber Security Strategy, and the eSafety Commissioner's recent request for industry to revise the Search Engine Code to cover developments in the field of generative AI and its integration into search engine functions.[13]

As AI is already subject to numerous laws and regulations in Australia, overseen by numerous regulators, Ministers, and Government Departments, an important first step in any future efforts to establish a regulatory framework for AI will be to consider domestic policy coordination. Inadequate coordination creates the potential for significant challenges for the Australian technology sector, and indeed, the economy more broadly. These include:

- Inhibited innovation and barriers to being able to capture the positive benefits of digital technology such as AI, resulting from inconsistent, unpredictable regulation;

- Unintended negative consequence of siloed technology policy development and narrow focus;

- Uncertainty and complexity for Australian businesses of all sizes grappling with changing and intersecting regulation; and

- Conflicting outcomes driven by inconsistent policy objectives and execution.

- Difficulties and complexities for Australian AI innovators to export globally

- Australia being excluded from international governance protocols such as the Hiroshima process agreed to be established by the G7 nations.

- Enhancing collaboration among the various policymakers and stakeholders involved in technology regulation in Australia would promote a more thoughtful, consistent, and effective development of AI governance.

(e)   Improving international coherence

In the borderless digital world in which AI operates, internationally coherent governance frameworks are key to achieving consistent, effective regulation that reflects foundational priorities and interests.

Bespoke local regimes can lead to misalignment with other jurisdictions that increases complexity and cost for businesses of all sizes. To move quickly on addressing the challenges raised by AI and capturing the benefits of this technology, it would make practical sense to leverage and align with international government and industry-led frameworks and standards – for example, the NIST's AI Risk Management Framework, which Microsoft has committed to implementing. We have also announced our support for the new White House Voluntary AI Commitments to help ensure that advanced AI

---

[13] eSafety Commissioner, Media Release, eSafety Commissioner makes final decision on world-first industry codes, 1 June 2023.

systems are safe, secure, and trustworthy.[14]

Aligning Australia's regulatory approach to AI with the work being done overseas and at the international level will increase legal certainty and lower regulatory burden, which should in turn generate economic growth by encouraging Australian and international companies alike to launch, grow, invest and remain in Australia.[15]

(i) Leverage and align with appropriate international standards

International standards are an effective vehicle for achieving and facilitating alignment in the globalised economy. They provide a consistent platform for innovation, safety and agreed behaviours that drive international trade and economic growth. Their utility is proven across the most complex of industry sectors, such as health, aviation and cybersecurity, both in domestic and cross-border settings. In the AI context – where we see a combination of complexity and rapidly advancing technology – standards play a critical role in clarifying best practice, procedures and societal expectations.

Microsoft views international standards as foundational to the consistency and interoperability of responsible AI. The international standards process also ensures all countries have a voice in setting these foundations. International standards carry the potential to assist compliance with regulations and support local industry (and the public sector) embed, and through recognised certification programmes, demonstrate responsible AI practices in their day-to-day operations.

International standards are also flexible regulatory tools that bridge the gap between government objectives and practical implementation. They can be reviewed and updated as global technology and best practices evolve – and can be implemented more swiftly than legislation. This flexibility is critical given the speed at which AI systems, and the associated challenges and opportunities, are developing.

AI standards can enable regulation to keep pace with broad consensus across stakeholder groups. Views on responsible use of AI will continue to develop through:

- International bodies such as the International Organization for Standardization (ISO), the International Electrotechnical Committee (IEC);

- Coalitions such as the OECD and Global Partnership on Artificial Intelligence (which bring together experts from science, industry, civil society, international organisations and government);

- Forums for open discussion such as the UK AI Summit in 2024 (the first major global summit on AI safety); and

- Domestic coalitions such as the National AI Centre (NAIC) established by the Commonwealth Government and led by the CSIRO.

There is significant opportunity to leverage international standards which emphasise good governance, risk management, and impact assessment. Key international standards under development include those being progressed by ISO/IEC.

---

[14] Microsoft on the Issues blog, Our commitments to advance safe, secure, and trustworthy AI, 21 July 2023.
[15] See Tech Policy Design Centre, Cultivating Coordination, 21 February 2023, p8.

> **ISO/IEC standards under development**
>
> **ISO/IEC 42001:** The flagship standard for certifying the AI management system within an organisation. It provides guidance, requirements, and controls for establishing, implementing, maintaining and continually improving an AI management system within the context of an organisation. Microsoft and many of our industry peers, customers, and suppliers, will be certifying our AI systems to this standard once it is finalised in the coming months.[16]
>
> **ISO/IEC 42006:** A standard which will establish the requirements for certification bodies to reliably audit and certify AI management systems. This standard will enable the assessment of the competencies of conformity assessment bodies in an efficient and harmonised way and ensures the comparability and reproducibility of certificates confirming conformity with ISO/IEC 42001.

Standards can also be woven into legislative responses to increase the flexibility of regulation and the ability for industry, government and academia to co-design appropriate requirements. For example, the EU's draft AI Act is intended to rely on harmonised standards (including those from ISO, IEC and European standards bodies) as a means of enforcement and compliance.

Microsoft supports exploring how international standards and conformance infrastructure can be further leveraged to respond to the opportunities and challenges posed by AI in Australia. We see this as a co-regulatory and multi-stakeholder process, with industry, technical experts (including Standards Australia) and relevant Australian governments being part of the conversation.

(ii) A voluntary international code of conduct

At the annual G7 Summit in Hiroshima in May 2023, leaders committed to "advance international discussions on inclusive artificial intelligence (AI) governance and interoperability to achieve our common vision and goal of trustworthy AI, in line with our shared democratic values."

Microsoft endorses the development of a voluntary, cross-border AI code of conduct. Technology development and the public interest will benefit from the creation of principle-level guardrails, even if initially they are non-binding. A non-binding code of conduct would underpin interoperability and coherence of AI governance, ensure that the policy approach to AI technology reflects common and shared values, and guide both public and private actors.

Microsoft believes an international code should:

- Build on the work already done at the OECD to develop AI Principles to promote the use of ai that is innovative and trustworthy and that respects human rights and democratic values;[17]

- Provide a means for regulated AI developers to attest to the safety of these systems against internationally agreed standards (such as the ISO/IEC standards under development); and

- Promote innovation and access by providing a means for mutual recognition of compliance and safety across borders.

---

[16] Under Microsoft's Responsible AI Standard, we have released a Responsible AI Impact Assessment Template to share what we have learned, invite feedback from others, and contribute to the discussion around building better norms and practices around AI.

[17] See OECD, AI Principles.

Additionally, an international code should incorporate 'safety brakes' for AI systems that control critical infrastructure (as discussed further in Section 2(f) below), licensing for highly capable foundation models and AI infrastructure obligations.

Consideration should also be given to establishing additional international regulatory agencies (for example, based on the International Civil Aviation Organization) with responsibilities relating to an international code, especially in relation to high-risk AI systems; for example, those that affect critical infrastructure.

Microsoft sees opportunities for Australian voices to join the current multilateral discussions on an international code of conduct. In particular, in addition to the G7 commitment in Hiroshima, Microsoft supports the joint efforts of European and US lawmakers to develop a voluntary AI Code of Conduct, as announced by the EU-US Trade & Tech Council (TTC) in May 2023.

> *"Now technology is accelerating to a completely different degree than what we've seen before...something needs to be done to get the most of this new technology... We're talking about technology that develops by the month so what we have concluded here at this TTC is that we should take an initiative to get as many other countries on board on an AI Code of Conduct for businesses voluntarily to sign up."*
>
> European Union Executive Vice-President Margrethe Vestager, speaking at a meeting of the TTC (May 2023)

(f) Targeted reform to implement effective safety brakes

As discussed above and acknowledged by the DISR in the Discussion Paper, many of Australia's existing laws and ongoing reforms are already responding to AI technology. Against this background, efforts will be best placed toward considering the application of existing laws to AI and for relevant regulators and public authorities to help raise industry and consumer awareness of the application of these laws in an AI context, such as via compliance guidelines.

For example, an area where Microsoft sees a potential gap when it comes to AI regulation globally is in relation to **AI systems that control critical infrastructure**. Whether it be energy supply grids, telecommunications networks or sanitation systems, AI is already taking a more active role in helping to administer assets of critical importance to both the community and national security.

In our recent whitepaper, *Governing AI: A Blueprint for the Future*, we emphasise the importance of effective safety brakes for certain advanced AI systems that control designated critical infrastructure. These safety brakes should allow humans to intervene and take control of an AI system, particularly where unexpected malfunction or misuse could have significant consequences. From our perspective, implementing safety brakes in AI systems is not just a technical challenge, but also a governance challenge. It requires clear rules, guidelines and training about when and how humans should intervene in an AI system's operation.

Australia is well placed to take on this challenge given it already has a comprehensive critical infrastructure framework. However, as such laws are primarily focused on matters of system security and cyber resilience, there may be gaps to fill in order to address more specific AI challenges that stem more from system design and the training of underlying models.

Additionally, the concept of safety brakes should be a key element of a voluntary, internationally-coordinated code of conduct as mentioned above.

Overall, where gaps in existing legal frameworks require legislative reform, Microsoft recommend that these be approached collaboratively with industry. As discussed further below in Section 3(c), there are significant opportunities for new public-private partnerships to work on addressing the challenges and benefits of AI. Microsoft recognises the work done in Australia already to identify and explore collaboration and coordination between the public and private sectors, including by the Australian National University's Tech Policy Design Centre and School of Cybernetics, the HTI, and the NAIC.

4 Supporting AI adoption in Australia

The responsible adoption of AI has the potential to drive significant economic growth and productivity in Australia, both now and into the future. As has already been experienced by many public and private sector deployers, AI can automate routine tasks, augment data, enhance decision-making, generate novel content, and create entirely new products and services. Not only do these capabilities deliver value directly to those entities, but also contributes to sectoral innovation, greater market competition and broader economic growth.[18]

The unique value of the generative AI of today (and what is yet to come) is the ability for businesses and institutions of all sizes to benefit from highly capable models that are pre-trained, continuously improving and increasingly intuitive to use. While insufficient stockpiles of training data, limited expertise and integration delays have been noted as barriers to effective uptake to date, advances in generative AI are set to dull the impact of these challenges.

As recognised by the Discussion Paper, lower relative AI uptake in Australia may hold the local economy back from taking full advantage of the AI opportunity. While Microsoft is confident that as AI improves and becomes more widely accessible, increased adoption rates will follow, this should be complemented by cooperation between industry and government to support local innovation and adoption of AI.

(a) Transparency as a generator of trust

Transparency, one of our six foundational principles discussed above, is a cornerstone for any discussion regarding AI. In our view, there is a great opportunity for transparency efforts to build trust in AI systems, and drive adoption, amongst the Australian public and local industries. Importantly, transparency is not just about explaining the technical complexities, but also about providing clear, understandable information to the public about the system's purpose, capabilities, limitations, and risks.

We have embraced transparency at Microsoft by implementing a range of measures, including transparency reports, user-friendly explanations of our AI systems, and tools that help users understand and control how their data is used and managed.

Based on our experience, ensuring transparency in relation to AI systems should involve a holistic approach that includes:

1   Ensuring that AI systems are designed to inform the public when they are interacting with an AI system (such as through user interface features, or labelling AI-generated content so that the

---

[18] For example, recent research from the Tech Council of Australia and Microsoft projects that generative AI could enable Australian workers to make productivity gains that have the potential to add $40-105 billion to the economy annually by 2030. New goods and services created by generative AI have the potential to contribute a further $5-10 billion. The range in potential value is based on the pace of adoption of generative AI technologies.

public understand the source of the content);

2   Clearly communicating an AI system's capabilities and limitations (including its risks), such as through documentation like our [Transparency Notes](#);

3   Ensuring AI models are 'explainable', which means they will be able to reasonably justify their decisions and how they reach their conclusions (and consequently, supports compliance with company policies, industry standards, and government regulations);[19]

4   Educational materials, such as our [Bing AI primer](#); and

5   Reporting on organisational policies, systems, progress and performance in managing AI responsibly. For example, transparency reports are a proven and effective way of driving corporate accountability, educating the public, and tracking progress on responsible AI practices. At Microsoft, we publish a variety of transparency reports and have committed to annually publishing an AI transparency report to inform the public about our policies, systems, progress, and performance in managing AI responsibly and safely.

Moreover, transparency is not a one-time effort, but a continuous process that involves regular updates and improvements based on user feedback and technological advancements.

Microsoft believes that government has an important role to play in promoting transparency in relation to AI, especially with regard to the deployment of advanced AI systems. Consideration could be given to initiatives for sharing information with the public on high-risk AI systems. For example, the [City of Amsterdam's Algorithm Register](#) is a public register which provides information on high-risk AI and measures undertaken to ensure safe and responsible use. Microsoft supports the creation of a national registry of high-risk AI systems. This registry should allow members of the public to review an overview of the system as deployed and the measures taken to ensure the safe and rights-respecting performance of the system.

The academic community also requires greater access to resources for critical research into the risks and benefits posed by AI and support to ensure public accountability, including through analysis of the behaviour of commercial AI. This includes support to meet the high cost of accessing computational resources. Further, based on our experience sharing our own cutting-edge foundation models with the academic community,[20] we believe that encouraging AI developers to share access to their technology with researchers will further enable the study of frontier applications and the sociotechnical implications of AI models.

Microsoft notes that transparency is not one-size-fits-all, and sometimes tensions exist between transparency and system security. With reference to the technology stack discussed at Section 2(b) above, it will be important to identify the level of transparency that is appropriate in particular circumstances or at particular levels. We suggest that any mandatory AI transparency measures be balanced with competing concerns such as confidentiality, privacy, cybersecurity and online safety.

---

[19] See Microsoft, 'Trust and understanding of AI Models' predictions through Customer Insights', 1 March 2022, available at: https://www.microsoft.com/en-us/research/group/dynamics-insights-apps-artificial-intelligence-machine-learning/articles/explainability/

[20] For example, through our [Turing Academic Program](#) and Accelerating Foundation Models Research Program.

The Frontier Model Forum

An example of our commitment to transparency is the recent announcement of the launch of the Frontier Model Forum, an industry body founded by Anthropic, OpenAI, Google, and Microsoft.[21] This Forum is focused on ensuring safe and responsible development of 'frontier AI models', defined as 'large-scale machine-learning models that exceed the capabilities currently present in the most advanced existing models'.

One of the core objectives of the Forum is to collaborate with policymakers and other important stakeholders to share knowledge about trust and safety risks. In the coming year, a key area of focus for the Forum will be to establish trusted, secure mechanisms for sharing information among companies, governments, and relevant stakeholders regarding AI safety and risks.

This forum will draw on the technical and operational expertise of its member companies to benefit the entire AI ecosystem, such as through advancing technical evaluations and benchmarks, and developing a public library of solutions to support industry best practices and standards.

(b)   Modernising regulation to address known blockers

Regulation can sometimes contribute to conditions where international investment and local innovation are stifled. In an AI context, it may be necessary or desirable to revise elements of Australia's existing laws that may undermine the ability of the local tech industry and consumers to take full advantage of emerging systems.

In Australia', some of this work is already started. For example, we welcome the ongoing work of the Commonwealth Government in **modernising Australia's privacy laws**. As noted in our response to the recent *Privacy Act Review Report*, Microsoft supports strong privacy laws that promote trust by ensuring responsible data use in line with emerging global norms. We are encouraged by the Government's consideration of ways to facilitate secure cross-border data flows, which is vital to facilitating the benefits of data-sharing across the economy while ensuring Australians' personal information remains protected regardless of its location.

In addition to privacy reforms, Microsoft sees a great opportunity for local **copyright laws** to be modernised to keep pace with digital change. There is a lack of clarity as to whether Australian copyright law permits **text and data mining (TDM)**. TDM refers to the use of automated techniques to analyse large volumes of data, identify patterns, concepts and produce new information. Such methods have been instrumental in the development of computer science, especially in relation to AI and large language models. Uncertainty surrounding use of TDM may exist under the copyright laws of some jurisdictions,, including Australia.[22] There is a widely held view among technology providers that the lack of clarity on TDM has become a blocker to homegrown innovation in AI.

While not squarely within the scope of the Attorney-General's Department's ongoing *Copyright Enforcement Review,* submissions from several tech providers and industry associations noted the continuing impediment that Australian copyright laws have on digital innovation. The introduction of clearer fair dealing exceptions, greater flexibility and alignment of safe harbour protections with international standards, were among the suggestions fielded. As AI becomes an increasingly routine part of how the Australian public interacts with and sources information, it will be important to preserve the community's ability to learn and derive knowledge from copyrighted works. We believe

---

[21] Microsoft On The Issues Blog, Microsoft, Anthropic, Google, and OpenAI launch Frontier Model Forum, 26 July 2023.
[22] See e.g. https://www.alrc.gov.au/publications/8-non-consumptive-use/text-and-data-mining.

this can be done in a way that respects the interests of rightsholders. Performing TDM on a work protected by copyright is not a copyright infringement. As required under Article 9(2) of the WTO Agreement on Trade-Related Aspects of Intellectual Property Rights,[23] copyright should not extend to ideas, procedures, mathematical concepts, etc. Overall, given the recent step change in AI, and the importance of TDM to AI systems, we recommend that Australia adopts an explicit fair dealing exception for TDM.

(c)   Public-private partnerships

Microsoft is keenly aware of the potential for AI to be used with varying levels of regard for the public good. Our work in responsible AI aims to promote the use of these advanced systems for good; as tools to protect democracy and fundamental rights, promote inclusive growth, and advance the planet's sustainability needs. We see significant opportunities to bring the public and private sectors together to work towards leveraging the potential of AI, and address the societal challenges associated with this technology. Increased trust and confidence in AI technology will flow naturally from public-private collaborations.

For example, there has recently been an increasing focus on the new risks to democracy and the public posed by applications of AI to alter content, such as creating 'deepfake' content. A collective effort is needed to combat threats like these. Microsoft is already actively taking steps with collaborators to tackle disinformation, including through the Coalition for Content Provenance and Authenticity, which was co-founded by companies including Microsoft, Adobe, the BBC, Intel, Sony and Truepic. We have also recently announced new media provenance capabilities coming to Microsoft Designer and Bing Image Creator in the coming months that will enable users to verify whether an image or video was generated by AI by using cryptographic methods to mark and sign AI-generated content with metadata about its origin.[24]

We are also investing deeply in Media Provenance capabilities, for example, by building tools into Microsoft Designer and Bing Image Creator by default such as embedded content tracing and provenance by default.

> "It's going to take a multi-pronged approach, including education aimed at media literacy, awareness, and vigilance, investments in quality journalism--with trusted reporters on the ground, locally, nationally, and internationally, and new kinds of regulations that make it unlawful to generate or manipulate digital content with an aim to deceive."
>
> Microsoft's Chief Scientific Officer, Eric Horvitz (31 January 2022)

(d)   Education and awareness

Ensuring Australian workers and businesses have the necessary skills and knowledge will be essential to capturing the opportunities and managing the risks of these technologies. Microsoft has long supported broad educational initiatives on new technologies, beginning in the 1990s during the mainstreaming of the personal computer. Increasing awareness and sharing information has always underpinned the adoption and realisation of the potential benefits of any new technology, and AI is no different.

---

[23] https://www.wto.org/english/tratop_e/trips_e/intel2b_e.htm#generalobligs
[24] https://blogs.microsoft.com/blog/2023/05/23/microsoft-build-brings-ai-tools-to-the-forefront-for-developers/.

We seek to lead by example by investing significantly in education relating to AI both internally and externally. In addition to focusing on upskilling within Microsoft, we also share our learnings, help others meet their regulatory requirements for responsible AI and support their implementation of responsible AI governance systems and practices. This approach is captured by our AI Customer Commitments.

> ## Microsoft's AI Customer Commitments
>
> - **We will share our learning about developing and deploying AI responsibly:** This includes sharing our AI knowledge and expertise with the public, including governments and our industry peers around the world, so that others can also learn from our experiences. We will also share the work we are doing to build a culture of responsible AI at Microsoft, such as parts of our internal training curriculum. Additionally, we will invest in dedicated resources and expertise to respond to questions about deploying and using AI responsibly. Ultimately, Microsoft believes in teaching by doing, and demonstrating what responsible AI practices look like at a best practice standard.
>
> - **We support the responsible implementation of AI systems:** We offer a dedicated team of AI legal and regulatory experts as a resource to support the implementation of responsible AI governance systems, and also work with partners to assist mutual customers in deploying their own responsible AI systems.

Microsoft believes there is opportunity for Government to support its own agencies and industry in taking a holistic, collaborative approach, which considers the whole regulatory landscape and reflects dialogue within and between the public and private sectors, and consumers. We note that the DISR and the Digital Transformation Agency have released interim guidance on government use of publicly available generative AI platforms, which is a useful first step in providing generative AI guidance to public sector staff.[25]

Additional opportunities to support the public sector in the use of generative AI could include initiatives to:

(a)   promote communities of knowledge sharing, which would enable collaborative learning through sharing and learning from each other's best practices, expertise and experiences;

(b)   assist businesses to navigate existing and upcoming regulation, and to understand their views on the effectiveness of the regulatory landscape;

(c)   explore and promote the benefits of AI, including potential applications and use cases;

(d)   raise awareness and educate consumers on AI regulation and risks;

(e)   empower consumers and stakeholders to share their views and experiences with AI, which would assist in identifying concerns, minimising unintended consequences, mitigating risks and ensuring AI design practices are inclusive;

(f)   learn from regulatory and industry experiences overseas; and

---

[25] https://architecture.digital.gov.au/guidance-generative-ai.

(g)   foster a culture of responsible AI approaches and practices in government agencies to serve as exemplars of what best practice looks like to the private sector, which could involve requiring government agencies to implement mandatory responsible AI controls, risk management frameworks, or standards.

5   Conclusion

Microsoft encourages a multi-stakeholder, collaborative and risk-based approach to AI governance in Australia. This would ensure informed and coherent outcomes across the different parts of the AI ecosystem and across the diverse sectors impacted by AI. Coherence across both domestic laws and international frameworks is essential to both effectively safeguard the Australian community as well as to ensure local industry can harness the enormous potential of AI.

We have always seen the progression of responsible AI as a journey. Microsoft welcomes the opportunity to engage with others on this journey: our industry peers, government, consumers, and other stakeholders. We look forward to exchanging insights and experiences on developing, using and regulating AI responsibly, including best practices and governance mechanisms, in the spirit of collaboration and mutual learning.

Attachment A - Commentary on Discussion Paper definitions

Discussion Paper definition	Microsoft commentary
Artificial intelligence (AI) refers to an engineered system that generates predictive outputs such as content, forecasts, recommendations or decisions for a given set of human-defined objectives or parameters without explicit programming. AI systems are designed to operate with varying levels of automation.	While this definition has indeed been taken from ISO/IEC 22989:2022, it draws primarily from the distinct concept of 'AI system' (3.1.4) and lacks the additional context provided along with that entry. These contextual notes acknowledge the use of various techniques and approaches to AI, as well as the role of AI in conducting tasks and operating with varying levels of automation. The ISO/IEC 22989:2022 definition of 'AI System' also relies on several other cascading definitions within the standard, and for that reason may be less effective when used in isolation.

The Discussion Paper definition also focuses on AI that 'generates predictive outputs' without mentioning the various other use cases of AI, including the execution of functions and representation of data.

We recommend that the OECD definition of AI, cited in the HTI's 'The State of AI Governance in Australia', provides both a more fulsome and flexible definition of AI:

Artificial intelligence ('AI') is a collective term for machine-based or digital systems that use machine or human-provided inputs to perform advanced tasks for a human-defined objective, such as producing predictions, advice, inferences, decisions, or generating content.

Some AI systems operate autonomously and can use machine learning to improve and learn from new data continuously. Other AI systems are designed to be subject to a 'human in the loop' who can approve or override the system's outputs. AI systems can be custom developed for a specific organisational purpose. Many are embedded in products or deployed by suppliers in upstream or outsourced services.

The NIST also provide a more straightforward definition that may better accommodate the breadth of AI:

Artificial intelligence (AI): (1) A branch of computer science devoted to developing data processing systems that performs functions normally associated with human intelligence, such as reasoning, learning, and self-improvement. (2) The capability of a device to perform functions that are normally associated with human intelligence such as reasoning, learning, and self-improvement. |
| *Machine learning are the patterns derived from training data using machine learning algorithms, which can be applied to new data for prediction or decision-making purposes.* | This definition should be reframed to place a greater emphasis on machine learning as a broad set of models and algorithmic techniques that enable machines to improve with experience. The current definition places too much emphasis on the output of such approaches.

We also note that, like 'artificial intelligence', 'machine learning' encompasses a range of techniques and can be considered an umbrella concept under which terms such as deep learning also sit. |
| *Generative AI models generate novel content such as text, images, audio and code in response to prompts.* | We suggest that the reference to 'in response to prompts' be removed or that the definition otherwise make clear that generative AI can also respond to inputs other than conventional user prompting. |

*A **large language model (LLM)** is a type of generative AI that specialises in the generation of human-like text.*	While predominantly known for their generation of human-like text, LLMs are also able to generate other forms of output such as code, images, video and the translation between these formats. We recommend that references to AI outputs here and across other definitions are not unduly prescriptive.
***Multimodal Foundation Model (MfM)** is a type of generative AI that can process and output multiple data types (e.g. text, images, audio).*	A broader definition of 'foundation models', under which MfMs, LLMs and other large-scale models fall, may be more appropriate for the scope of the Discussion Paper. The following is drawn from the research of Stanford University's *Center for Research on Foundation Models:* *Foundation models are highly capable AI models trained on large datasets at scale that can be adapted and applied to a wide range of downstream tasks.*