FORHUMANITY

Response to Supporting Responsible AI

Australian Government call from the Department of Industry, Science and Resources

Contributors:

Ryan Carrier, Executive Director
Elle Brooker, ForHumanity Fellow
Diana DiCecco, ForHumanity Contributor
Fernando Mourão, ForHumanity Contributor

Contact Information

Ryan@forhumanity.center
https://forhumanity.center

Address: 980 Broadway #506
Thornwood, NY 10594, United States

Summary

Enclosed please find a submission from ForHumanity responding to the Australian Government's 'Supporting Responsible AI in Australia" discussion paper from our Team Australia comprising more than 40 members from across Australia.

The submission:

- responds to the government's 20 key questions in detail,
- proposes that the Australian and State Governments consider taking a new, proactive approach to regulating new services and products and minimising harm by mandating 'responsibility by-design' and assurance by third-party mandatory independent audits
- requires product and service developers and controllers to assess and be responsible for the impacts of what they take to market, regardless of how it is constituted; and
- flags our groundbreaking work creating a global new industry, modelled on financial audit, for independently certifying organisations in a standardised way for which we seek your government's input and accreditation.
- Agrees that "*Many private and public organisations are already adopting ethical principles or similar practices to ensure appropriate accountability and governance mechanisms are in place for AI. These include:*
    - *major tech firms such as Microsoft, Google, Salesforce and IBM*
    - *public organisations such as the US Department of Defense and the Australian Signals Directorate.*

About ForHumanity

ForHumanity has been developing Independent Audit of AI Systems for more than six years via a crowdsourced and transparent community of global experts, specialising in AI, Algorithmic, and Autonomous (AAA) Systems. For Australia, Currently ForHumanity is drafting criteria to enable third party independent audits of AAA Systems that would meet

the requirements identified in both the CSIRO Ethical Framework and the NSW government responsible AI frameworks.

#	Australian Government Question	ForHumanity Response
1.	Do you agree with the definitions in this discussion paper? If not, what definitions do you prefer and why?	"This paper uses the key definitions below, which are based on the International Organisation for Standardization (ISO) definitions" For more detailed technical definitions, see the ISO's definition of terms related to AI (ISO/IEC 22989:2022). —------- The current definitions have a number of technical inaccuracies and imitations. For instance: 1. "Predictive outputs" (we also have descriptive outputs, for example: automatic summarisation of documents) 2. "Machine learning are the patterns derived from training data" (machine learning are not patterns, and are not restricted to training data - for instance clustering) ForHumanity prefers the OECD definitions as they are more technically accurate, and their adoption would enhance international consistency and by extension Australia's economic competitiveness and capacity to seamlessly trade with EU member states. In its landmark law - the EU AI Act. The European Union has cooperated with the OECD to agree upon a definition of Artificial Intelligence. The result is likely to impact any definition that Australia establishes as Australia looks to harmise under OECD and related international treaties. —--

#	Australian Government Question	ForHumanity Response
		—----------
		As observed by the authors who use " *A broad definition of AI … that includes any products or services using AI techniques…. rang(ing) from simple rules-based algorithms guided by human-defined parameters to more advanced applications like neural networks"* Australians and local media are using 'AI' as a blanket term for things that it meaningfully doesn't include, and in ways that lack nuance and clarity concerning different technologies and their use cases, and particular their context based risks.
		As an example, in the paper "Bias can occur when datasets used to train a model or algorithm are not comprehensive" pg 8
		However, bias can also occur at the design level, (as well as we be 'by design', as happened with Robodebt).
		The misuse of a broad definition has enabled citizens and the media to 'distinguish' Robodebt from otherwise positive and open responses to 'AI' when in fact, the paper's definition expressly includes it. The definition is a misdirection, and as such, misinformation.
		It is the use cases, the context, the application and the risk arising which are the salient features making these applications and instruments ethical or responsible, or not, and why.
		As a result, we advocate that you adopt a different term, so that Australian citizens have a better understanding of the nuances and what's at stake.
		At ForHumanity we use the term "AAA systems" to include and distinguish between algorithms (such as Robodebt), automation (for example machine learning and services such as ATO eTax) and true AI (artificial intelligence) to refer to a wide range of emerging instruments and services needing mindful consideration about their impacts and benefits.
		We would prefer the Australian Government to take a more nuanced approach to these concepts, and to

#	Australian Government Question	ForHumanity Response
		understand that asking Australians about responsible AI is as helpful as asking them about 'responsible transport', 'responsible gambling' or 'responsible gardening'. "AAA systems" are more meaningful.
2.	What potential risks from AI are not covered by Australia's existing regulatory approaches? Do you have suggestions for possible regulatory action to mitigate these risks?	A. General harms arising from foreseeable risks and fast to market / Agile products predicated on the existing regulatory approach of reacting to proven harm = potential and long standing risk to Australia's existing regulatory approach(for example: A framework that allowed the tobacco industry to thrive until it could be medically and ethically proven to cause cancer). B. Yes, we do. As you observe *"... all technologies, (including) emerging technologies such as AI bring new opportunities but also new challenges."*[1] Why not regulate based on the harm a product or service can do, or foreseeably will do, and embed responsible design and harm minimisation into the fabric of every new tool, technology, service or offering regardless of its scale, scope, sector or industry? —---------- Australia's existing regulatory model is a patchwork quilt of reactive responses to issues of the day, that reflect both the distribution of power between the different levels of government, and the political compromises needing to be reached to get ideas across the line at the relevant time.

[1] page 7

#	Australian Government Question	ForHumanity Response
		Respectfully, this quilt, and the holes in it, are a risk in and of themselves. For example: Regulating privacy based on organisational size has exposed Australian citizens to avoidable harm, that the proposed update closes the gate on after the horse has bolted. Similarly, enabling an employment exemption of any kind, rather than regulating the data regardless of the context for collection, makes little sense. A retinal scan is more identifying than a fingerprint Appreciating that there is a division and distribution of power and responsibility between the levels of government, but also, that we live in an era aspiring to citizen centred government - can the Australian and State governments contemplate a shift away from reactions to specific problems, almost alwys after the fact of a harm, (for example every new drug or technology being met with its own regulatory response) and into the realm of eliminating risks and threats to humans, harm reduction by-design and excessive penalties for use cases that lead to harms to people, setting citizen protection as a baseline, using risk methods? The UK Children's Code is perhaps the world's first example of an attempt to regulate products and services on a sliding scale based on the user's age and capacity. Shifting responsibility for preventing harm in the first place onto all industries would redefine how business is done at every level and across all sectors and promote the mindset, infrastructure, accountability 'by design' and risk capability and reduce the regulatory burden that organisations so often complain about. If the burden was as simple as 'first, do no harm' - that's a principle that it's difficult to see any sector complaining is expensive and best avoided.

#	Australian Government Question	ForHumanity Response
3.	Are there any further non-regulatory initiatives the Australian Government could implement to support responsible AI practices in Australia? Please describe these and their benefits or impacts.	ForHumanity is in the process of creating a new industry, modelled on financial audit,for independently evaluating and certifying AAA systems deployment and development taking a jurisdictionally-centric and human centred approach to developing criteria, but seeking government approval when applicable. Our hope is that the State and Australian governments will endorse, and formally accredit the framework for use on shore as part of an independent, viable, holistic and sustainable approach to regulation with the capacity to assess and validate compliance with the law, that can be applied evenly and equally to government departments, universities, developers, deployers and to industry, by third party independent auditors. The benefit to Australian citizens is the cross-jurisdictional compliance and high rights protection without the investment or tax burden of yet another new Commission, or the problem to fitting AI into an existing jurisdiction ill equipped to handle AAA systems and their impacts. A bit like the Heart Foundation's Health Tick the scheme operates on an industry funded, opt in basis that in turn allows consumers and citizens to have some agency over their exposure to AI via the presence or absence of a ForHumanity certificate.
4.	Do you have suggestions on coordination of AI governance across government? Please outline the goals that any coordination mechanisms could achieve and how they	ForHumanity provides a top-down governance, oversight, and accountability structure that is applicable across all sectors and use cases of AI and is globally harmonised for Australia's international organisations. In addition, our bottom-up approach allows the Australia government to examine each individual use case (for example, we are developing audit criteria for 82 different use cases of Automated Employment Decision Tools) Independent Audit of AI Systems enables: • Interoperability, particularly in eHealth. • Harm minimisation and prevention • External, independent and ongoing certification and validation

#	Australian Government Question	ForHumanity Response
	could influence the development and uptake of AI in Australia.	• Citizen agency • Data Sovereignty - by which we mean that the laws of Australia apply to data that belongs to Australian citizens regardless of where it is on-shored 'in the cloud'. • Simplicity* • Understandability • Human-centricity and support of fundamental human rights • Sustainability, Ethical Social Impact and good governance (ESG) • Consistency with international standards and relevant legal frameworks without these needing to be in place. There are enough international sign posts that ESG and stakeholder capitalism are becoming dominant that Australians must learn to lead by example when to do so is in the public interest. *It's worth noting that 'getting to simple' is rarely simple. Simple outcomes, and easy to use systems, are the product of complex processes. The perceived lack of alignment between process and outcome, or blind faith in cause and effect, leading to cognitive dissonance has perverse impacts on the quality of inputs and research into 'what works'. As an example, the production of two pages of advice for the Minister in other jurisdictions is an executive summary of a larger piece of work that also goes to the Minister. Only in Australia has this executive summary assumed the status of a full briefing, with adverse impact (as evidenced at the Robodebt Royal Commission.)

#	Australian Government Question	ForHumanity Response
5.	Are there any governance measures being taken or considered by other countries (including any not discussed in this paper) that are relevant, adaptable and desirable for Australia?	Yes. **EU AI Act, Digital Services Act, General Data Protection Regulation, UK Children's Code** **US Blueprint for an AI Bill of Rights** https://www.whitehouse.gov/ostp/ai-bill-of-rights/applying-the-blueprint-for-an-ai-bill-of-rights/ **The OECD Framework for digital talent and skills in the public sector** https://www.oecd-ilibrary.org/governance/the-oecd-framework-for-digital-talent-and-skills-in-the-public-sector_4e7c3f58-en ForHumanity, has an agenda of seeking government endorsement and accreditation wherever possible, from regulators such as the Australian Government in support of the rule of law, as a sign that they meet local and relevant regulatory standards and requirements. In consultation with the UK Information Commissioner Office (ICO) and the European Data Protection Board, For Humanity is in the process of seeking government endorsement of our certification schemes. We also advise, under contract, the European standards body CEN/CENELEC, who is charged with producing conformity assessments under the EU AI Act.
6.	Should different approaches apply to public and private	No with exceptions, such as military applications, intelligence, and criminal investigations.

#	Australian Government Question	ForHumanity Response
	sector use of AI technologies? If so, how should the approaches differ?	ForHumanity's approach is to maximise harmony with audit criteria that drive governance, oversight, and accountability for all AAA Systems. We then use a bottom-up approach for each individual use-case of AAA Systems to identify unique and specific risk. What is unique and tailored to each individual AAA System is the risk of harm or negative impact to humans. Therefore, a risk based approach is merited.
7.	How can the Australian Government further support responsible AI practices in its own agencies?	Establish a public framework of applicability equally to both public and private applications. Require mandatory certification for transparency and third party affirmation of compliance with the framework
8.	In what circumstances are generic solutions to the risks of AI most valuable? And in what circumstances are technology-specific solutions	In our opinion, the "... *circumstances (in which…) generic solutions to the risks of AI most valuable…?"* are when the focus of them is on reducing or eliminating harms to humans, upholding their rights, the rule of law, and their democracies, *by design*. There is some support in the paper for this argument ***"By-design considerations***

#	Australian Government Question	ForHumanity Response
	better? Please provide some examples.	*These are becoming increasingly popular as preventative mechanisms to ensure the design of appropriate AI or other digital systems. They include privacy by design, data protection by design (DPbD) and safety by design. They ensure AI systems are designed with privacy or safety considerations in mind from the outset.* *For example, DPbD allows digital systems to automatically delete data once it is no longer needed for the specific business purpose. These design concepts can be voluntary or mandated through laws." pg 31* —-- Automation and machine replacement of labour has been a productivity driver across economies for generations. AI is not meaningfully different except for the white collar jobs that are likely to be supplanted or required to become more expert and specialist, in accounting, medicine and law. There are multiple use cases demonstrating that repetitive, teachable, high impact tasks, such as scanning medical images for anomalous cells in which the results are escalated to a human for further investigation are immensely valuable. We would argue that this does not make these tools meaningfully "intelligent" any more than automation has meant this in other industries.
9.	Given the importance of transparency across the AI	

#	Australian Government Question	ForHumanity Response
	lifecycle, please share your thoughts on: a. where and when transparency will be most critical and valuable to mitigate potential AI risks and to improve public trust and confidence in AI? b. mandating transparency requirements across the private and public sectors, including how these requirements could be implemented.	The Disclosure of Residual Risk, transparently to users and AI Subjects to enable them to be informed users of the tools. Furthermore, ForHumanity considered each risk control, treatment and mitigation for merit of transparency. Data Transparency Documents, like Model cards and Datasheets for DataSet, Privacy Policies, Automated-decision making explainability statements are all examples of transparency that lead to robust feedback loops for the public and the companies that produce AAA Systems. However, complete transparency, as is sometimes required by regulatory oversight, does not support innovation, protect intellectual property and trade secrets. That is the intersection where Independent Audit of AI Systems balances the benefits of transparency with the benefits to society of opacity. Trust is earned in many ways, providing transparency, providing explanations, disclosing Residual Risk, providing mechanism to remediate problems, such as Adverse Incident Reporting Systems. However, Independent Audit of AI Systems is the most powerful tool to build an infrastructure of trust as demonstrated with a 50+ years track record of financial audits. Knowing that a third-party will independently assure compliance changes behaviour and leads to dramatically increased compliance.
10.	Do you have suggestions for:	

#	Australian Government Question	ForHumanity Response
	a. Whether any high-risk AI applications or technologies should be banned completely? b. Criteria or requirements to identify AI applications or technologies that should be banned, and in which contexts?	ForHumanity recommends that Australia identify and maintain a list of Prohibited AAA Systems that do not proportionally uphold protection for rights and freedoms of individuals and the country's shared moral framework (including but not limited to, social behaviour scoring, emotional recognition systems without construct validity and ground truth validation). —---------- Consistent with ISO Standards, and accepted best practice in risk management, High Risk is a function of context and use case, or likelihood x impact, not the fact (or future fact) of a technology, in and of itself. Technology is disruptive, rather than moral or capable in this way, and the evidence shows that banning technologies completely tends to have absurd effects, unless the harms are known or easily foreseen, and even then there will be risk takers willing to experiment on themselves in ways that no Ethics committee would ever approve of [2] The leading law internationally, on how to regulate on a risk basis, (specifically, how user age impacts on permitted and prohibited use, and how to build this into the design and delivery of software services and systems) is the UK Children's Code[3] Regulating technology, (or for that matter, all humanity and industry regardless of its status), according to the foreseeable harm that it can do would be a novel method for assessing what needs immediate regulation, and what already falls into a prohibited category. However, the novelty of any such legal framework, coupled with

[2] https://www.vice.com/en/article/avy3m4/an-afternoon-with-some-melbourne-biohackers
[3] https://ico.org.uk/for-the-public/the-children-s-code-what-is-it/

#	Australian Government Question	ForHumanity Response
		Australia's broad disinterest in leading the pack internationally in digital in this way, the way it once did[4] not to forget the need for parliaments and departments to be seen to have ongoing work to do which perhaps explains the case by case approach to new technology means that this proposal is unlikely to be accepted, trialled or tested on Australian shores drawing on the evidence. This recommendation conflicts with the arguments put forward in the report to the effect of" *"These are areas where the government has deemed specific sector-specific laws are necessary. Sector-specific regulations need to be well designed to avoid duplicating economy-wide regulations while filling in any gaps appropriate to AI." pg 10* and *"As an enabling technology, AI is increasingly combined with other components and emerging technologies to produce innovative new businesses, products and services. This often means that AI is regulated under multiple laws, increasing the likelihood of possible duplication or conflict between regulatory systems, and associated compliance burdens on AI developers and adopters." pg 13* Respectfully, we disagree with this view and the necessity for it because of the legal conflicts it invites. To ForHumanity, this seems like a misuse of thinking, resources and time designed to generate avoidable conflict and effort.

[4] Victoria became the first jurisdiction in the world to appoint a Minister for 'Multimedia' in 1999, and also passed the globally unique Health Records Act in 2002. Since that time New Zealand, the UK, the EU and the US have all moved further forward, faster and more professionally than Australia has, until the COVID-19 crisis created the momentum former NSW State Minister Victor Dominello needed to cut through red tape with the exception of the appointment of an eSafety Commissioner Julie Inman, an American.

#	Australian Government Question	ForHumanity Response
11.	What initiatives or government action can increase public trust in AI deployment to encourage more people to use AI?	This seems like an appropriate point to endorse "**Box 3: Australia's AI Ethics Principles** 1. **Human, societal and environmental wellbeing:** AI systems should benefit individuals, society and the environment. 2. **Human-centred values:** AI systems should respect human rights, diversity, and the autonomy of individuals. 3. **Fairness:** AI systems should be inclusive and accessible, and should not involve or result in unfair discrimination against individuals, communities or groups. 4. **Privacy protection and security:** AI systems should respect and uphold privacy rights and data protection, and ensure the security of data. 5. **Reliability and safety:** AI systems should reliably operate in accordance with their intended purpose. 6. **Transparency and explainability:** There should be transparency and responsible disclosure so people can understand when they are being significantly impacted by AI, and can find out when an AI system is engaging with them. 7. **Contestability:** When an AI system significantly impacts a person, community, group or environment, there should be a timely process to allow people to challenge the use or outcomes of the AI system. 8. **Accountability:** People responsible for the different phases of the AI system lifecycle should be identifiable

#	Australian Government Question	ForHumanity Response
		and accountable for the outcomes of the AI systems, and human oversight of AI systems should be enabled." pg 14 ForHumanity makes 10 explicit recommendations: **RECOMMENDATION #1** ForHumanity recommends that the Australian Government establish oversight and control of the production and usage of AAA Systems to ensure the safe usage of these tools for all individuals, especially for vulnerable populations. Successful and robust oversight can be found in markets such as: 1) Drug Testing, clinical trials, and usage/labelling approvals and disclosure of residual risk in the form of side effects 2) Safety testing of airlines and automobiles prior to approved usage 3) Oversight of model risk management in financial companies overseen by regulators (SR 11-7) **RECOMMENDATION #2** ForHumanity recommends to the Australian Government that tit establish an oversight body comprising the relevant Federal Departments and Agencies to oversee Certifying Bodies and to approve Audit Criteria.

#	Australian Government Question	ForHumanity Response
		RECOMMENDATION #3 ForHumanity recommends that Parliament identify and maintain a list of Prohibited AAA Systems that do not proportionally uphold protection for rights and freedoms of individuals and the country's shared moral framework (including but not limited to, social behaviour scoring, emotional recognition systems without construct validity and ground truth validation). **RECOMMENDATION #4** ForHumanity recommends the adoption of a risk-based approach to the governance, oversight, and accountability of AAA Systems that includes Diverse Inputs and Multi-Stakeholder feedback to identify and mitigate harms and negative impacts to individuals. Better risk assessment processes will result in a clearer delineation of risky and non-risky applications of AAA Systems that allow for: 1) Focused use of resources on high risk applications of AAA Systems and 2) Reduced regulatory burden on small and medium enterprises where risks to individuals are limited

#	Australian Government Question	ForHumanity Response
		RECOMMENDATION #5 ForHumanity recommends that all AAA Systems that are subject to a mandatory, annual, risk-based audit conducted by an independent third party using a robust definition of "Independent" that ensures the elimination of conflicts of interest unless they are: 1) Prohibited or 2) Deemed to be low risk This system establishes a national network of oversight built into business operational compliance that proactively ensures compliance with laws and regulation and does not rely upon reactive Federal enforcement. **RECOMMENDATION #6** ForHumanity recommend that the Oversight Body described in recommendation #2, ensures that all entities providing certification are accredited initially and reviewed and monitored regularly to uphold the following characteristics: 1) They are liable for false assurance of compliance 2) They are qualified to provide expert level service 3) They are held to a standard of Professionalism and Code of Ethics 4) They have robust systems to support integrity and confidentiality of Independent Audits

#	Australian Government Question	ForHumanity Response
		5) They have robust systems, processes and procedures to uphold Independence **RECOMMENDATION #7** ForHumanity recommends a dual approach to certification criteria: 1) Top-down governance, oversight and accountability that is applicable to all AAA Systems 2) Bottom up, sector-specific and use case specific criteria tailored explicitly to mitigate risk in the sector and use case **RECOMMENDATION #8** ForHumanity recommends that Audit criteria cover the following areas of concern: 1) Legal and Regulatory compliance 2) Ethical oversight, proportional protection for rights and freedoms of individuals and the management of tensions and trade-offs amongst priorities 3) Data Management and Governance, including assurance of Data and Information Quality 4) Fairness-by-design, including active bias mitigation and avoidance of unfair and deceptive practices 5) Privacy-by-design and Data Protection 6) Technical documentation, including User Guides for downstream user transparency 7) Risk Management process that allows for the assessment of a collection of identified risks, risk indicators and emergent risk 8) Monitoring, including post-market, continuous and adverse incident reporting systems

#	Australian Government Question	ForHumanity Response
		9) Human oversight and human control 10) Cybersecurity minimum standards 11) Upstream/downstream a) Liability (resolving the conflict between zero product liability (software) and strict product liability (automobiles)) b) Supply chain certainty and transparency through independent assurance of supplier satisfaction of downstream accountability c) Assurance of provenance and upholding intellectual property rights, such as copyright with traceability and chain-of-custody procedures **RECOMMENDATION #9** ForHumanity recommends that the audit criteria are representative of risk to humans and drafted in a transparent, accessible, inclusive manner that maximises compliance harmony across sectors and use cases. **RECOMMENDATION #10** ForHumanity recommends that the criteria, standards and benchmarks achieve "auditability". Often, laws, guidelines, and standards have insufficient specificity to achieve this goal which results in subjective compliance that short-circuits one of the great values of Independent Audit, which is uniformity of compliance compared across many sectors and use cases. Auditability means "binary" (compliant/non-compliant) criteria where an at-risk auditor is able to determine compliance or non-compliance with certainty.

#	Australian Government Question	ForHumanity Response
		Evaluation methods for assurance will include public and private documentation, logs, registers, and databases, official filings, pictures and graphics and physical testing to affirm compliance. All audit compliance should be retained for a period that is consistent with the law and that ensures a meaningful lookback period to ascertain errors in the chain of production.
12.	How would banning high-risk activities (like social scoring or facial recognition technology in certain circumstances) impact Australia's tech sector and our trade and exports with other countries?	Minimally - Australia, in Parliament, will have to determine the balance between protections of fundamental rights of individuals and the broad community shared moral framework. ForHumanity supports the democratic process that establishes the choices made in support of that shared moral framework.
13.	What changes (if any) to Australian conformity infrastructure might be	Establish government bodies that can adjudicate certification scheme criteria to ensure that binary (compliant/non-compliant) audit criteria reflect the intent of the laws and regulatory frameworks. Then encourage a robust, uniform ecosystem of third party independent audit and the requirement for AAA Systems

#	Australian Government Question	ForHumanity Response
	required to support assurance processes to mitigate against potential AI risks?	to be compliant to be used in government and private markets
14.	Do you support a risk-based approach for addressing potential AI risks? If not, is there a better approach?	Yes. However, 'risk based' starts with robust assessment by a pool of diverse human assessors and multi stakeholder feedback designed to consider risk from a 360 degree perspective. Each individual AAA System will have different levels of risk to humans. Compliance requirements should be measured as a function of those potential impacts and associated mitigations.
15.	What do you see as the main benefits or limitations of a risk-based approach? How can any limitations be overcome?	Risk is oriented in perspective. A new cancer treatment that could cure a dying person is valued completely differently than the same drug taken by someone who is not dying. Therefore, risk-based approaches require evaluation by a 360 degree perspective of diverse inputs and multi stakeholder feedback. 1. AI Risk levels are not homogeneous at a sector/domain level: It is simplistic and unrealistic to assume that all AI services from a single sector have the same risk level and/or tolerance. 2. AI Risk levels are not static: AI models are adaptive and transient artefacts, and model drifts are inevitable for the sake of the accuracy and efficiency of AI systems. 3. AI Risk classifications are heuristic decisions based on partial, contextual, and transient understanding: Setting a rigid classification schema for AI risks based on our current knowledge will create an unstable and floating system that would

#	Australian Government Question	ForHumanity Response
		introduce uncertainty and discontinuity of interpretation over time. Therefore, ForHumanity advises a human-centric Algorithmic Risk Assessment and Ethical Risk Assessment, including the use of Fundamental rights, data protection, and systemic societal impact assessments as forms of risk-based assessment and the establishment of risk controls, treatments, and mitigations leading to a transparent disclosure of Residual Risk to end users
16.	Is a risk-based approach better suited to some sectors, AI applications or organisations than others based on organisation size, AI maturity and resources?	Unlike most other management tools and frameworks, risk is intuitive, scalable, and transcends industries, sectors, maturity and technologies because of its adaptability and capacity to introduce legal as well as human centred factors: both workforce (internal) and stakeholders (external). We would argue that it's suited to any and all organisations and many other human endeavours besides. A domain-specific framework, that is ISO aligned and that dynamically reassigns risk priorities for AI systems based on the use cases, number and nature of potential risks, contextual information about deployment and data drifts, and responsible AI principles and specific norms more concerning for each use case. A risk-based approach struggles with new and emerging technology (which is why ForHumanity requires all novel technologies to be treated as high-risk until some maturity is achieved)
17.	What elements should be in a risk-based approach for addressing potential AI risks?	As an example, we are submitting excerpts of our GDPR Certification scheme and Children's Code certification scheme, as samples

#	Australian Government Question	ForHumanity Response
	Do you support the elements presented in Attachment C?	
18.	How can an AI risk-based approach be incorporated into existing assessment frameworks (like privacy) or risk management processes to streamline and reduce potential duplication?	ForHumanity's risk management framework is fully integrated and operationalised ISO, NIST, COSO risk management processes
19.	How might a risk-based approach apply to general purpose AI systems, such as large language models (LLMs) or multimodal foundation models (MFMs)?	Independent Audit of AI Systems is a coordinated upstream/downstream delineation of responsibilities and accountabilities for providers, deployers and users of AAA Systems, including foundation models (LLM/LMMs) Our certification schemes ensure that upstream providers and enabling compliance for downstream deployers including terms of service, disclosures, transparency and most notable Residual Risk. Essentially, all parties know "who does what to whom" upstream and downstream. A contributor with expertise in ethical datasets put it best when she described LLMs as "being in their

#	Australian Government Question	ForHumanity Response
		Napster phase and needing to move to the Spotify model"[5]. For more than six years ForHumanity has been drafting audit criteria that enables organisations to assess their compliance with globally significant legislative developments as well as their ethical performance, with transparency, explainability and certified ongoing monitoring and responsiveness. Risk is at the centre of this approach. Regardless of what one prefers to think about LLMs and generative AI, and their potential, the principle "garbage in, garbage out" still applies to all data driven endeavours and most organisations have thoroughly underestimated the quality of their data, data governance and control, and underestimated the work involved to clarify the data the way you would a consomme, or butter, in a recipe to produce a clear and clearly higher grade, result. There is scope for individual organisations and academia to develop private LLMs, based on preferred and preferably the 'best' of human traits and knowledge, instead of all publicly scrapeable information, and it's these future LLMs that we would prefer to hold out for and to see how their biases and outputs are are more able to reify, evaluate, compare and contrast between. The role of government is in mitigating harms and upholding copyright with regard to LLMs, and protecting the vulnerable. Segments of our community regard creativity as a human trait that deserves protection from efforts to erode it for cost-saving, dishonest, misleading and exploitative reasons, because of the new economic potential and value it adds that is tangible.

[5] Keren Flevell,

#	Australian Government Question	ForHumanity Response
20.	Should a risk-based approach for responsible AI be a voluntary or self-regulation tool or be mandated through regulation? And should it apply to: a. public or private organisations or both? b. developers or deployers or both?	Australia is a mixed market economy and "both" is the right response. Regulatory responses are almost always lag responses, responding to demonstrated harm. In addition to being blunt instruments, and drafted in a way that seems reluctant, or else unable to anticipate new developments, even when these are foreseeable, their role is limited and reactive and the process for passing legislation is slow and drawn out. Whilst self-regulation has been the model preferred by successive governments in Australia and elsewhere, as well as being the natural default for any new responsible and ethical roll out, AAA systems are conspicuous for the way in which they are attracting, open calls for government intervention from experts across the ideological spectrum, from Joy Bulumwini to Gary Marcus and Geoffrey Hinton. Coupled with developments in the UK and EU (GDPR, the AI Act and Children's Code) the demand for authorities to do something still comes with the caveat that neither legislation nor self-regulation imposes any impetus on an organisation to take decisions and actions, and have the infrastructure in place for ensuring, that its choices are demonstrably, conscienscientiously, with knowledge, explainability and transparency, demonstrably ethical and responsible. The results of any number of Royal Commissions at State and national levels, highlight that industries and elite leadership figures in all sectors, health, family justice, banking and superannuation, aged care and Human Services, lag behind community and legal requirements and expectations and that something new is needed. ForHumanity is proposing a new industry in independent, opt-in audit of AAA systems, modelled on Financial audit, that is trans-national, self-funding, voluntary, as well as free of the conflicts that the PWC scandal (and arguably Robodebt in which Accenture seem to be conspicuously absent in taking responsibility for their role) have manifested by being profit motivated.

#	Australian Government Question	ForHumanity Response
		ForHumanity is a not for profit organisation working with regulators to uphold the Rule of Law and the democratic principle of citizen delegated power at the centre of it in ways that possibly haven't had traction since the founding fathers in the US litigated its constitution, the UK moved from being a sovereign controlled power to a parliamentary democracy or Australia federated. —------ B. The question aligns to approaches in the EU and the UK and for this reason we support the effort to ensure Australia is taking a globally consistent All organisations using data tend to use, adapt, and adopt the same tools to store, govern, and dispose of the information, meaning it is irrelevant to the impacts and harms who it was and what sector they were in that collected the data in the first place. Regulating government separately to the private sector gives the private sector tacit permission not to have to be respectful or mindful of rights and community standards although the Robodebt scandal has levelled the playing field in being "venal". There is another option that is more cost effective and less onerous, as well as more practical and likely to meet Australian Citizens' needs and that is: not-for-profit.

About us

ForHumanity (https://forhumanity.center/) is a 501(c)(3) non-profit organisation dedicated to addressing the Ethics, Bias, Privacy, Trust, and Cybersecurity in artificial intelligence, algorithmic and autonomous (AAA) systems.

Mission

ForHumanity's mission is to:

> *"....examine and analyse downside risk associated with the ubiquitous advance of AI, algorithmic and autonomous systems and where possible to engage in risk mitigation to maximise the benefits of these systems… ForHumanity*

ForHumanity's certification schemes:

- Address specific local legal requirements
- Draw on relevant trans-national, socio-technical and industrial standards and human rights-based risk approaches and
- Incorporate international best practice, making our approach valuable and useful to organisations in ways that stand-alone national schemes can never be.

When supported by independent pre-audit services, (gearing organisations up to meet the scheme's requirements), as well as a method for systematically and continuously reviewing and verifying compliance in an ongoing manner, citizens and regulators can be confident that the ethical deployment of AI and algorithms is robust.

Our Approach

ForHumanity outputs have been crowdsourced using an open and transparent process that draws from a pool of over 1400+ international expert and interested contributors from more than 85 countries to construct accessible and actionable:

- audit criteria,
- certification schemes, and
- educational programs for legal and compliance professionals, educators, auditors, developers, and legislators to aid them in mitigating bias, enhancing ethics, protecting privacy, building trust, improving cybersecurity, and driving accountability and transparency in AI, algorithmic and autonomous (Triple AAA) systems .

ForHumanity works to make AI safe for everyone and makes itself available to support government agencies, regulators, entities and instrumentalities to manage the socio-technical and human rights risks associated with AI, algorithmic and autonomous systems.

We know that AAA Systems, using Personal Data, have been placed on the market with insufficient governance, oversight and accountability, including failures of technical, ethical, regulatory and organisation controls, resulting in harm to citizens.

To address this, ForHumanity has developed a systematic risk mitigation process, to ensure that future harms and failures are avoided, mitigated, treated or prevented, and that privacy and data protection are maximised.

The system is called the ***Independent Audit of AI Systems (IAAIS).***

The aim of the scheme is to promote and operationalise compliance with the law and international best practice by entities opting in to the scheme, and to license, professionalise and empower new certification bodies and accredit practitioners with the expertise to complete independent audits that assess and certify compliance, with assurance, on behalf of the public.

Why this scheme?

Unlike processes/systems in the past, and simple, stand alone, tools, such as a calculator or desktop computer, AAA Systems are **socio-technical**, which means that:

- humans, and the reading, interpretation and motivated application of their private, anonymised, Sensitive and Personally Identifying Data, via their apps, devices and detectable routines, and
- our humanity defined as our rights, freedoms, equality, and dignity (as well as how those rights and entitlements are interpreted, exercised, applied and experienced) are integral to the function, purpose, operation and intended performance of these tools, processes and systems.

The convergence, between AAA System advancements, Australian's cyber rights and the desire for convenience, efficiency and profit, creates unique, specialised and multi-disciplinary challenges to the safe and beneficial operation of these AAA systems in the interests of humans by organisations.

The normative criteria in this certification scheme have been specifically tailored to foster conformity with AI Assurance Frameworks from New South Wales[6] and the CSIRO's AI Ethics Framework[7], in an implementable way, that can be assured by independent auditors in consultation with Australian-based experts.

The scheme:

- Represents the practical and reasonable steps that Providers can follow and rely on as indicators and sign posts of good governance, control and implementation of AAA systems by operational teams, with oversight from Ethics, Risk, Internal Audit, users, executives and boards

[6] https://www.digital.nsw.gov.au/policy/artificial-intelligence/nsw-artificial-intelligence-assurance-framework
[7] https://www.csiro.au/en/research/technology-space/ai/AI-Ethics-Framework

- Encourages globally consistent best practice and greater compliance with international law and human rights based expectations, in a way that's independently and publicly assured 'at arm's length' from management, with traceability and accountability, for the benefit of all Australians, their rights, and their trust in institutions and the Rule of Law
- Offers independently validated and certified recognition that an organisation deploying AAA systems has sufficient and useful risk mitigations in place to offset, prevent, or otherwise minimise known socio-technical, human and technological downside risks including:

Conformity with these frameworks and this certification scheme means that Providers have maximised risk mitigations associated with socio-technical systems present, including:

a. Unfairness - statistical, cognitive and non-response bias in data and systems
b. Unfairness - failure to apply Algorithm Ethics to the data process
c. Non-Transparency - lack of explainability / 'black box' deployment
d. Lack of accountability - Insufficient governance and oversight, balances and checks.
e. Insufficient integrity and security from new vectors of cyber attack
f. Purpose limitation -
g. ModelDrift (including data and concept drift) - as a result of machine learning in automated decision making, and profiling that deviates from scope, nature, context and purpose.

Around the world there have been calls and claims made for internally developed, self-assessed ratings and conformity assessments, intended to steer organisations towards ethical and high performing governance, oversight and accountability of AAA systems, including in Australia.

Self-assessment, and self-regulation, is the typical 'first-step response' to the legal, economic and humanitarian challenges that disruptive technologies such as AI, Algorithmic and Autonomous (AAA) systems represent to existing and accepted ways of operating.

Organisations around the world have voiced concern about the financial burden of regulation[8], whilst simultaneously claiming their own unself-critical ability to provide trustworthy AI or to have created an internal and unvalidated framework that is unique to them.

[8] To different effects, see Adam Leon Smith discussing regulatory impacts as catalysts for innovation https://asmtech.com/podcast/adam-leon-smith/

Robodebt, which has been found to have been venal, incompetent and to require enabling legislation, demonstrates most organisations will default to self-interest or profit over human rights, public interest and compliance, even in the Government sector without independent oversight.

Self-assessment is an important first-step forward, but can only be the endgame for systems and processes that are low-risk on the human-rights scale. Any mature system, in which meaningful risks are present, will eventually demand third-party, independent certification for certain accountability in order to properly manage risk, segregate duties, assign liability, demonstrate trustworthiness and draft contracts.

ForHumanity enables an infrastructure of trust, modelled on existing financial audit approaches, in support of the inevitable maturity in governance, oversight and accountability for an AAA System[9].

An infrastructure of trust, as it relates to certification, is an unconflicted process, deploying a segregation of duties, conducted by certified and trained experts, that establishes a robust ecosystem that engenders trust for all citizens, regulators and Australian stakeholders and protects those who have no power or control.

Infrastructure of Trust

For Humanity's system is grounded on four core tenets:

1. ForHumanity produces accessible, binary (compliant / not compliant) certification criteria that transparently and inclusively aligns to Public requirements, in the EU, (e.g. GDPR, EU AI Act) that embeds compliance and performance into practice, and is considerate of corporate wisdom, but impervious to corporate dilution and undue influence, while being mindful of the regulatory burden and dedicated to maximising risk mitigations to humans (ideally criteria is approved and mandated by governments or regulatory bodies)

[9] In Australia, the media, academics and industry have taken to using 'AI' as a byword for everything from machine learning to ordinary computer programming, to language models such as ChatGPT, BlockChain and crypto minting, transformation or digital automation, as well as artificially intelligent, autonomous and algorithmic systems. 'AAA systems' is the more accurate, explainable, understandable and accepted concept internationally and is what Australians conventionally mean when they refer to 'AI' in writing and conversation.

2. Individuals are trained and accredited on certification criteria as experts. They are individually held to a high standard of behaviour and professionalism as described in the ForHumanity Code of Ethics and Professional Conduct - they are ForHumanity Certified Auditors (FHCAs)

3. Certification Bodies employ FHCAs to independently assure compliance with certification criteria on behalf of the public. They are licensed, independent, robust organisations that take on the task and risk, on behalf of the public to ascertain assurance of compliance. They are held to standards of independence and anti-collusion and are further subject to third-party oversight ("watching the watchers"), by entities such as national accreditation bodies (e.g. COFRAC, DaKKE).

4. Corporations use the criteria to operationalise governance, oversight and accountability for their AAA System that helps them to satisfy certification compliance comprehensively. Comprehensive compliance will create leverageable governance, oversight and accountability that will simultaneously lead to more sustainable profitability and reduce the risk of negative outcomes for their stakeholders

Laws are reactive and designed to encourage compliance, but they do not assure it. Independent audits performed by third parties which apply criteria required by statute, regulation, or industry custom, such as the Generally Accepted Accounting Principles (GAAP) or International Financial Reporting Standards (IFRS), create a system that encourages proactive compliance. ForHumanity believes this system represents a more trustworthy environment for the processing of personal data using AI, autonomous systems and emerging technologies.For the purposes of certification, an 'infrastructure of trust'