CONCORDIAVOX

+61 435 948 464
james@concordiavox.au
212/132 Smith, St
Collingwood, Victoria, 3066

To:

Technology Strategy Branch
Department of Industry, Science and Resources
Industry House, 10 Binara Street
Canberra ACT 2601

26 July 2023

Dear General Manager,

Thank you for the opportunity to contribute to the Australian Government's consultation on additional AI governance mechanisms.

This submission contains a brief overview of Concordia Vox and our interest this consultation process as well as responses to select questions posed by the discussion paper.

We have also prepared three overarching recommendations that consolidate these responses into practical policy or governance measures and are designed to complement many of the initiatives already identified within the paper. These recommendations are to:

1. Extend the Consumer Data Right to cover AI

   The discussion paper provides a comprehensive list of principles, standards, codes, and laws that might govern AI in Australia. Also important are the administrative or technical mechanisms necessary to give effect to these various frameworks, to ensure the barriers to consumers for asserting their rights aren't so prohibitive as to render them meaningless (nor so cumbersome for AI vendors and buyers to stifle their use in Australia).

   Australia's Consumer Data Right (CDR) is a good example of how a well-designed administrative mechanism can give effect to policy intention in the digital era. Although the CDR was designed to help consumers share data between providers in search of a 'better offer', it could conceivably be adapted to facilitate some of the disclosure and consent rights flagged in the discussion paper (e.g. requests from consumers for details of when AI is being applied to their data).

   While the Australian Government has paused the sector-based roll out of the CDR beyond energy and banking, a separate implementation track could be developed for progressing AI-related data rights. To allow for a rapid implementation, AI-related disclosures could start as voluntary in the first instance (offering responsible AI vendors and buyers a means of differentiating themselves in the market).

2. Develop an Employee Data Right (EDR)

   The *Privacy Act Review*, referenced in the discussion paper, highlighted how private sector employee records are exempt from existing protections, meaning employees are starting farther

behind than consumers when it comes to managing their interests as AI is deployed. Assuming the Government proceeds with the recommended consultation on extending privacy protections to employees, parallel work could commence on the design of an Employee Data Right to give administrative effect to those protections as well as any that arise from this process.

3. Establish a right to digital representation

The final and perhaps most substantial recommendation is for the Australian Government to support representative voices in AI policy and delivery by establishing a right to 'digital representation' for end users, in particular employees and consumers. Such a right would allow end users to nominate a third party to exercise certain individual rights on their behalf (e.g. request access to training data or to revoke consent) and/or impose as a 'negative right' on AI vendors or buyers that would prevent them from refusing to recognise representations made digitally on behalf of their end users.

Enabling end users to engage with AI vendors and buyers via a digital representations made by third parties would bring the following benefits:

- Streamlined accountability arrangements to the benefit of all participants. For example, AI vendors and buyers can deal with bulk representations made digitally rather than submitted one-by-one from individual end users.
- Increased community trust and confidence in AI governance and delivery from having representative voices involved in relevant decisions (enabling more ambitious innovations to be pursued).
- The option of providing confidential disclosures to representatives for circumstances where there might be legitimate concerns regarding disclosure to end users (or what is may effectively be public disclosure).
- Increased awareness of individual rights leading to their being exercise earlier in the AI lifecycle, ensuring regulators and courts can focus more on systemic issues.

We'd be pleased to expand on any of the information provided in this submission or respond to any additional questions you may have.

Yours faithfully,

James Pawluk
Managing Director

About Concordia Vox and its interest in AI

Concordia was founded in 2022 to help companies and their investors maximise long term returns by ensuring alignment with the interests of their customers and workforce. In effect, we are tackling the first incarnation of AI's 'alignment problem', with notable AI experts[1] describing corporations as a form of artificial intelligence that is already more intelligent than any individual human being. While 'corporate intelligence' cannot continuously improve at a faster rate than its underlying human intelligence, making alignment easier to achieve over the long term but susceptible to short-term lags of misalignment which is what our work seeks to help overcome.

From an operational perspective, artificial intelligence and machine learning form a substantial part of our research and operations (upwards of 75 per cent of costs to date) and in the process we've acquired an understanding of the effort required to train sufficiently explainable models to support effective decision-making. By opting to focus on applying AI technology, rather than building it, we are maintaining visibility of emerging language-based AI tools and how vendors are promoting the various applications and benefits to their target customers (not all of which are is always aligned with the interests of end consumers and employees).

One overarching insight to emphasise for this consultation goes to the unpredictable nature of large 'black-box models'[2] that even their creators acknowledge. So long as these models remain a black-box, with interpretability and explainability unsolved, demonstrable alignment of inputs is paramount as misaligned inputs (i.e. training data and intentions/applications) will almost with certainty give rise to misaligned AI models. An onus should be on AI vendors (and potentially major buyers) to substantiate alignment of both their training data and the use cases of their customer base to provide confidence that unexplainable models are not being skewed from the outset. As the current investment hype settles and AI risk is integrated into ESG investing, as with climate risk before it, we can expect investor demands for increased stakeholder alignment (covering customers and workforce at a minimum) for companies want the trust to push the boundaries when it comes to developing or deploying AI[3].

[1] See comments from Max Tegmark in interviews with Lex Fridman and Robert Wiblin & Keiran Harris
[2] See Pause Giant AI Experiments: An Open Letter
[3] See the Norway's sovereign wealth fund's plan to issue guidelines on ethical AI

Response to questions

1. *Do you agree with the definitions in this discussion paper? If not, what definitions do you prefer and why?*

   The definitions in Figure 1 are appropriate though would benefit from the inclusion of definitions for artificial general intelligence (AGI) and artificial superintelligence (ASI) given the elevated risks some AI safety experts forewarn of should those thresholds ever be passed (e.g. the possibility that an ASI might kill all of humanity). On the other hand, there are also experts that view such alarmist predictions as a diversion from tackling the significant risks posed by existing AI capabilities[4]. Defining these terms could assist to clarify the scope of this process and focus the energy and attention of contributors.

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?*
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.*

   There are two inherent risks that cannot be addressed by existing approaches (i.e. those summarised in Figure 3 of the discussion paper):

   - Risk of regulators lagging innovators. This first risk is often floated in the context of new technologies and goes to the accumulation of knowledge and expertise in the hands the companies at the forefront of innovation, moving at a pace that is impossible for legislators and regulators to keep up. Perpetual asymmetries in information and resources imply that regulations will be at best ineffective (as companies with a 'disruptive' mindset find ways around them) or at worst stifle all innovation putting the jurisdictions that introduce them at an economic disadvantage.
   - Risk of regulatory capture. This second risk is also a well-established concept in economics and relates to the possibility that companies with malicious intentions may exploit the forementioned asymmetries to influence the design of regulations in their favour. This could include attempts to stifle competition by pushing for 'safeguards' that happen to create barriers for new entrants or watering down consumer protections by over-stating their costs or impacts.

   For clarity, while we judge these risks as material and warranting consideration as part of this consultation process, they don't provide justification for regulatory inaction (as some have argued). Instead, the best antidote to both risks ensuring strong and well-resourced stakeholder voices that are sufficiently independent of both industry and government.

   Therefore, the main **non-regulatory initiative** that the Australian Government should implement (question 3) is to ensure there is adequate space for such representative voices and to foster them through its own use of AI technologies (i.e. engage with public sector unions on the use of AI in relation to their work and engage with representatives of users of government services on the use of AI on its deployment in service delivery).

   In terms of **regulatory action** (question 2) the Australian Government should seek to support representative voices by establishing a right to 'digital representation' for end users, in particular employees and consumers. Such a right would allow end users to nominate a third party to

[4] See Statement from the listed authors of Stochastic Parrots on the "AI pause" letter

exercise certain individual rights on their behalf (e.g. request access to training data or to revoke consent) and/or impose as a 'negative right' on AI vendors or buyers that would prevent them from refusing to recognise representations made digitally on behalf of their end users.

A useful precedent of this is the role of unions in advancing improved occupational health & safety outcomes in individual workplaces and across the labour market more broadly (via contributing to public policy development). This has not only improved levels of safety but also increased level of trust in labour markets and their governing institutions, which in turn facilitates improved risk-taking by companies and allows them to take on more ambitious goals that contribute to the prosperity of the county.

In an AI context, jurisdictions with more effective digital representation will be able to facilitate higher levels of trust that unlocks increased consumer and employee support for deploying the technology, including on high value applications that using highly sensitive data.

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?*

   For mid-sized economies like Australia, the economic opportunities available from AI will likely be correlated with the level technological integration with larger markets. This means Australia inevitably will need to strike a balance between merely being a regulation 'taker' that maximises the level of integration with larger markets and crafting its own policies to suit local values and preserve Australian sovereignty. Indeed, some major AI vendors are already arguing for Australia to take a conformist approach.[5]

   This balancing act becomes particularly important if Australia wants to not only become a leader in responsible AI but to convert leadership into economic opportunities. With this goal in mind the Australian Government should consider where it can purposefully differentiate on governance measures to be able to attract 'preferred' capital that is being mobilised in the current investment boom (i.e. investment targeting responsible AI to generate long-term returns rather than capital swept up in the short-term hype).

   One simplistic option, the minimalist regulatory path, strives to compete with other countries for investment by promising industry minimal red tape and a low cost of doing business. But for mid-sized economies, outsized success won't just depend on attracting capital but attracting end-users that generate export earnings and this is likely to eventuate if a country is perceived as having relatively weak safeguards from an end user perspective.

   Instead of entertaining a false trade-off between safeguards and innovation, the emphasis should be on strong safeguards that minimise compliance costs (and enforcement) by designing administrative mechanisms that are streamlined and efficient while leveraging Australia's reputation for stable rule of law underpinned by non-partisan courts.

6. *How can the Australian Government further support responsible AI practices in its own agencies?*

   The Australian Government should lead by example in facilitating digital representation for both public sector employees as well as end users of government services.

7. *In what circumstances are generic solutions to the risks of AI most valuable? And in what circumstances are technology-specific solutions better? Please provide some examples.*

[5] See 'Microsoft urges soft approach as Husic vows to regulate 'high risk' AI'

This will partly come down to economies of scale and the challenge that any accountability mechanism will have in matching the capability of commercially deployed AI. The probability that effective mechanisms and solutions can achieve adequate capability (and scale to sustain it) increases when they can expand their scope across multiple technologies or sectors. Generic solutions, where feasible, should therefore always be preferred over technology-specific (or sector-specific) solutions.

An example of this is the recommendation for expanding the CDR to cover AI, a common data standards and in some cases infrastructure for facilitating the various consumer rights regarding AI will help to reduce costs for everyone involved by eliminating duplication.

11. *What initiatives or government action can increase public trust in AI deployment to encourage more people to use AI?*

    Once again this is where digital representation can play an important role. Public trust can be enhanced through awareness that well-resourced stakeholder groups representing their interests have been heavily involved in developing AI governance and that those same groups have the capacity to exercise various accountability mechanisms on their behalf.

13. *What changes (if any) to Australian conformity infrastructure might be required to support assurance processes to mitigate against potential AI risks?*

    The recommendation to expand the Consumer Data Right (CDR) to incorporate AI transparency will entail upgrading the associated infrastructure and services to support the data messages required to give effect to any new AI related rights (e.g. those relating to data requests or managing consent). Similarly, accreditation of data holders and recipients will need to be adapted to incorporate the organisations that act as an AI vendor/buyer or an end user representative.