



Submission to the Australian Human Rights Commission – Human Rights and Technology Discussion Paper

Introduction

The Commonwealth Bank of Australia Group (**CBA**) is Australia's leading financial services organisation and is recognised as one of Australia's most innovative companies.¹ We welcome the opportunity to respond to the Australian Human Rights Commission's Human Rights and Technology Discussion Paper (the **Discussion Paper**).

While the Discussion Paper does address emerging technologies more broadly, the focus of this submission will be AI-related issues.

With appropriate standards and practices in place, we believe that algorithms can deliver better outcomes compared to purely human decision making processes.² To this end, as a key contributor to the Department of Industry's Ethical AI Principles, and one of 5 national participants trialling their implementation, CBA is highly supportive of efforts to develop a common set of tools and guidelines to ensure that the promise of AI is realised in a safe, fair, transparent and ethical fashion, and have developed pioneering capability in this area. CBA fully supports the Commission's focus on the ethical development, deployment and use of AI and advocates for steps that encourage ongoing innovation and investment by Australian companies, as well as ensuring that AI delivers improved outcomes for consumers.

Understandably, like in other jurisdictions, there is an important role for statutory bodies such as the Australian Human Rights Commission to set out proposals in regard to the realisation and maximisation of the benefits of these technologies, and the elimination or minimisation of any potentially negative consequences. The Discussion Paper facilitates public discussion about these issues.

¹ The Australian Financial Review's 50 Most Innovative Companies list for 2016, 2017 and 2018.

² See note 4 above.

Proposal 1 – National strategy on new and emerging technologies

CBA supports the development of a *National Strategy on New and Emerging Technologies*.³ This strategy should contemplate existing national roadmaps, standards or initiatives issued by Government, agencies or other stakeholders, including:

- *AI Ethics Framework* released by the Department of Industry, Science, Energy and Resources;
- *AI Roadmap*⁴ and the *AI Ethics Framework*⁵ issued by CSIRO's Data61; and
- *Final Report – An Artificial Intelligence Standards Roadmap: Making Australia's Voice Heard* by Standards Australia.⁶

The Artificial Intelligence Standards are being developed concurrently with Australia's participation in international forums such as ISO/IEC JTC 1 (IT-043) (International Joint Technical Committee for ITC Standards which includes AI).

In developing a strategic approach to regulating AI, the Discussion Paper recognises the importance of leveraging existing laws, co-regulatory and self-regulatory frameworks that support the ethical use of AI, including human rights, anti-discrimination and privacy laws:⁷

"[t]he complex challenges presented by new technologies require a multi-faceted approach. This starts with applying current laws more effectively, and identifying gaps in the law that require reform. Where the law is appropriately silent or does not provide a clear rule to apply, other measures, such as ethical frameworks, can guide 'good decision' making"

We agree that the focus of Proposal 1 should be on applying existing laws to ensure that AI is used appropriately, including laws that:

- govern privacy;
- relate to human rights, including discrimination;
- provide remedies for negligent conduct;
- require services to be delivered fairly;
- require a service to be fit for purpose;
- require services to be delivered with reasonable care and skill;
- provide consumer guarantees on products and services; and
- prohibit misleading or deceptive conduct.⁸

Law reform in an AI context should be considered if there a manifest gap in existing laws and frameworks. Given the array of laws that currently implicitly or explicitly govern the use of AI systems, we expect that it would be valuable from a governance perspective to consider how these laws are best applied to optimise AI outcomes in conjunction with considering how standards can support or complement these laws.

³ While the focus of our submissions is on AI, CBA also considers that there is merit in extending such a strategy beyond AI to explicitly include 5G, AgTech, Internet of Things and Quantum Computing.

⁴ <https://data61.csiro.au/en/Our-Research/Our-Work/AI-Roadmap>

⁵ <https://data61.csiro.au/en/Our-Research/Our-Work/AI-Framework>

⁶ https://www.standards.org.au/getmedia/ede81912-55a2-4d8e-849f-9844993c3b9d/O_1515-An-Artificial-Intelligence-Standards-Roadmap-soft_1.pdf.aspx

⁷ Discussion Paper, 55.

⁸ In addition to these laws, there are a range of other sector specific regulation that implicitly or explicitly regulates the use of AI. For example, for financial services organisations, there is prudential regulation such as CPS 220 Risk Management and CPS 510 Governance and the Design and Distribution Obligations in the Corporations Act.

It is essential to support trials to help develop better standards and practice. CBA is part of an industry pilot with NAB, Telstra, Microsoft and Flamingo AI trialling eight ethical principles relating to AI, developed as part of the Government's AI Ethics Framework. The outcomes of this trial will 'deliver practical benefits and translate into real world solutions'.⁹

Any law reform that may ultimately be required should also be technology neutral. Regulating a decision-making process based on its technological form, and leaving other decision-making processes unregulated could have an unintended consequence. It may create a two-tier system delineating between AI users and non AI processes which may result in reduced levels of investment and adoption of AI in Australia. This would reduce the benefits that can be delivered to customers through innovation.

Proposal 2 - Ethical framework

CBA supports an ethical framework of the type contemplated by Proposal 2 and notes that the ethical considerations are being contemplated by the development of the standards mentioned in the preceding section.

Stakeholders such as the AHRC could consider becoming a nominating organisation within committees responsible for the development of these standards (eg, ISO/IEC JTC 1 (IT-043)) to ensure its objectives are met.

Question A - AI informed decision making

The Commission has advanced the following definition for discussion purposes:

AI informed decision making ...[means] a decision that has a legal, or similarly significant, effect for an individual and AI must have materially assisted in the process of making the decision.

This definition of 'AI informed decision making' has three key elements.

Defining AI

The first element is the term 'AI' itself. The definition of the term 'AI' is a keenly debated topic across industry, academia and government. The term refers to a wide range of technologies, including natural language processing, machine learning and deep learning which in turn can be implemented using a collection of processes, algorithms and systems. As AI is a general purpose technology, a general purpose forum such as the applicable standards body mentioned in respect to Proposal 1 above (i.e. ISO/IEC JTC 1 (IT-043) committee) would be an ideal forum for developing and publishing a definition of AI. It is an ideal forum as it will lead to a definition that can be used globally. We should avoid developing a definition of AI domestically that is inconsistent with a global standard. The forum, comprising experts from across academia, government and industry, is well-placed to develop a balanced definition of AI for general purpose use. Once published, the definition would help promote consistency and certainty across industry, academia and government.

⁹ <https://www.minister.industry.gov.au/ministers/karenandrews/media-releases/businesses-ready-test-ai-ethics-principles>

Legal effect

The second element ‘a decision that has a legal, or similarly significant, effect for an individual’ would cover a wide range of decisions. For example, entering into a contract of any type has a legal effect on a consumer so this element itself would not play a particularly strong role in how the definition would be applied.

Materially assisted

Finally, the third element, ‘AI must have materially assisted in the process of making the decision’ would likely capture many AI systems and models that are utilised in business and government today. For example, this broad definition would include any AI model output which is used in a wider process in which a decision is made. Such an outcome would increase implementation requirements for industry and government without any clear public benefit. The impacts of a wide definition would also influence the actions required to be taken in relation to other proposals in the Discussion Paper, including for example Proposal 5 and Proposal 8.

Proposal 5 - Informing people about the use of AI

Following on from our comments in relation to Question A, the definition of terms within the AHRC’s meta-definition should be refined to ensure its scope is appropriate as it would influence the notifications required to be provided under Proposal 5.

In addressing the notification issue, the Commission should also consider potential overlap with notifications required under privacy legislation (eg, regarding the use of data).

Further, we would recommend consideration be given to the potentially negative consumer impact if the requirement is cast too broadly. AI is increasingly used to provide improved customer experiences. For example, in terms of security of access to services (biometrics). If examples of this type were classified as materially assisting a decision and necessitated such notifications, the eventual ubiquity of the technology may lead to very poor user experiences and act to limit access to services intended to be beneficial. Further, notifications under anti-money laundering legislation could be unlawful and in other cases such as fraud detection would be imprudent to deploy as they may tip off bad actors.

Proposal 7 - Explainability legislation

We support the qualification in Proposal 7 being that an explanation would be required to be provided where an individual ‘...would have been entitled to an explanation of the decision were it not made using AI’. If this is a reference to legally entitled,¹⁰ this qualifier would strike an appropriate balance. It would also ensure that the approach was technology neutral.

We agree that an individual should be informed of a decision, as well as being given a succinct explanation of the reasons for the decision except to the extent it could disclose sensitive information such as:

- proprietary or sensitive information of government or industry; or

¹⁰ For example under s132 of the *National Consumer Credit Protection Act 2009* (Cth) a credit provider must provide a copy of a credit assessment upon request by a consumer and under s21P of the *Privacy Act 1988* (Cth) a credit provider must notify an individual if it has declined credit based wholly or partly on certain information from a credit reporting body.

- information in relation to algorithms implemented for anti-money laundering / terrorism financing detection, information security and fraud prevention.

In order to build trust in AI models, this right should be supported by effective governance principles and standards that would guide the appropriate risk assessments (including privacy impact assessments), notifications, design, model selection, use, supervision and monitoring (including periodic testing), complaint handling, auditing and accountability in respect of AI models.

We do consider that it is desirable to provide a technical explanation at large in addition to a lay explanation of the decision for the following reasons.

Complexity of AI models

AI models themselves are complex. In all cases an individual would require domain or business knowledge and then knowledge about the technical implementation of the model to be able to form a view as to how an outcome was generated. The review would also need to consider the legitimate trade-offs made between accuracy and explainability when selecting a model and the reasons for that decision. For instance, logistic regression is highly explainable but has low predictive accuracy because it is a simple model. On the other hand, very high dimensional deep learning algorithms applied to, say fraud detection data sets, can be demonstrated to be highly accurate but explainability for humans beyond identification of the largest contributory features may not be achievable in the conventional sense. This should not prevent any entity from harnessing the proven accuracy associated with these systems. For example, in improving the accuracy of voice recognition.

Personal, proprietary and other commercially sensitive information

The proposal could necessitate close evaluation of the inner workings of an AI model. This would involve the review of confidential information or intellectual property in many cases. It could also expose personal information to the reviewer. In providing a general right to explanations in the context of matters like credit risk decisions would lead to possible opportunities to reverse engineer credit decisioning logic or other commercially sensitive information.

The review process

The standard to which the person with relevant expertise would need to be satisfied is unclear. This uncertainty could impact the timeliness of a review.¹¹

In order to make a meaningful assessment of models, especially ones consuming large amounts of data, a review would need to contemplate the steps taken to avoid bias, discrimination/reconstruction, overfitting, data drift or model leakage. They would need to understand the approach to training, validation and hold out data sets, and the reasons for selecting the particular AI model deployed. They may seek to perform their own testing to support their observations. Particularly sophisticated AI models may also interact with real-time data streams and/or require very compute-

¹¹ The scope for divergent expert views about the features that impact an algorithm's output were evident in *Australian Competition and Consumer Commission v Trivago N.V.* [2020] FCA 16, at [106]ff.

heavy and time consuming training runs, further impacting the practicality of conducting multiple such reviews.

These steps could consume an inordinate amount of time and resources and this impact would be exponentially compounded if every person affected by a decision were able to make a request of this type. Of course, technical explanations of an algorithm and related decision making processes would still be required for oversight purposes, including regulators, dispute resolution bodies and courts.

Human in the loop/Human on the loop

Where there is a human on the loop (design and monitoring oversight) and a decision involves a human in the loop (capability for human intervention during the decision cycle) we see less justification for requiring a technical explanation if an appropriate governance approach applying common standards is applied to AI models.

We are not aware of any other jurisdiction where there is a specific right of technical explainability.¹²

We support transparency through the provision of lay or functional explanations supported by the ongoing development of standards we mention above and the industry trial in which we are participating. The standards would set out what is considered good industry practice which industry and government could adopt through the design, use and decommissioning lifecycle for AI models. In addition, industry trials could also serve as a vehicle for testing how some of the issues mentioned in this section could be addressed.

Proposal 8 - Non deployment on non-explainable models

We support the concept outlined in Proposal 8 if the relevant systems covered by Proposal 8 are those to which the qualifier in Proposal 7 applies (i.e. Proposal 7 would only apply were an individual would have been entitled to an explanation of the decision were it not made using AI). The person making the assessment in this case should be the person implementing the system on behalf of the organisation. Organisations should also consider how frequently these assessments should occur and in what circumstances these assessment should be subject to independent review by either an internal or external person. Such assessments would be covered by the good governance practices and standards we refer to above.

Proposal 9 – Centre of Expertise

We support this proposal. We recommend that the Centre focus on explainability should target emerging analytical techniques (i.e. unsupervised machine learning and neural networks and explainability techniques such as LIME or Local Interpretable Model-Agnostic Explanations). We should also recommend that the Centre focuses on the AI focus areas outlined within CSIRO's AI roadmap for Australia (i.e. industry sectors of interest).

¹² We note that Article 71 of the General Data Protection Regulation refers to a non-binding right of explanation in very general terms.

Question B – Proposed explainability presumption¹³

Introducing a presumption of the kind described would not be technology neutral in that it would treat decisions involving an AI input differently to purely human-based decisions.

The proposal would also be redundant where there are existing comprehensive legislative requirements relating to decision making processes. For example, the *National Consumer Credit Protection Act 2009* (Cth) sets out a range of requirements in respect to lenders engaging in credit assessments and related decision making processes.¹⁴ Given the highly prescriptive nature of those requirements, we consider that any new law of the kind described would not be necessary as a consumer's interests are well catered for in the design of the existing law.

Proposal 10 - Accountability for AI

Accountability within any organisations should align with existing legal and regulatory frameworks that apply to those organisations,¹⁵ including apportionment of liability principles.

Question C - Lawfulness of AI

In our view, it is essential to focus on internationally standardised framework for assessing the explainability and outcomes of AI, instead of introducing any new, bespoke mechanisms. See also comments under Proposal 10.

Proposal 11 - Legal moratorium on facial recognition

Decisions on how organisations adopt new and emerging technologies within their businesses should adhere to all required legal and regulatory obligations, alongside any internal cyber, privacy and risk management frameworks. The decision to adopt such technologies should also be pro-customer and focused on providing value to customer experiences.

Biometrics, as a family of technologies, offers CBA the ability to offer customers greater levels of security and convenience when accessing and offering our services. To date, our adoption of biometrics has focused on enabling the use of TouchID and FaceID through the CommBank App. These technologies do make decisions about who can access the App. However, they are not an exhaustive means for accessing our services (i.e. customers may also login to our App with their digital banking login and password/pin). We do not think it is the intention of the Commission that such use cases would fall within the scope of the proposed moratorium.

If CBA were to adopt further biometric solutions within our systems, these would be in line with our legal and regulatory obligations, our internal cyber, privacy and risk management frameworks, and be non-exclusive (i.e. other means for accessing our services would still exist and be available).

¹³ In responding to this question, we have presumed that this question relates to AI assisted decisions where an individual is entitled to an explanation under existing law (see comments under Proposal 7 above).

¹⁴ See ss128-132 of the *National Consumer Credit Protection Act 2009* (Cth).

¹⁵ For example, the Banking Executive Accountability Regime for Authorised Deposit-taking Institutions (which will be extended under the Financial Accountability Regime to all APRA regulated entities), Prudential Standards and existing laws already address accountability issues in a financial services context.

CBA agrees that it is important to ensure the use of facial recognition technology and biometrics is lawful, is protected from a cybersecurity and privacy perspective and that individuals consent to the information being collected and the uses specified.

As security, privacy and resilience of technology increases, CBA considers there will be a role for the expansion of biometrics such as the use of facial recognition technology noting the successful implementation of other biometric security measures such as touch, voice and iris identification and the use of other biometric data in applications such as health and fitness. The technologies can be implemented in a manner which is secure and protects the rights of individuals and also enhances the security levels generally in the environments in which they are used, which is pro consumer. Any proposal that directly or inadvertently impacts use of these technologies may disadvantage Australian companies from international competitiveness as well as limit the access of Australian consumers to improved services and security.

Question D - Intervention of human decision makers

Human decision makers are ultimately accountable for decisions that they, their teams, or technologies they deploy, make in the course of operating an organisation they are responsible for. See also our comment on accountability under Proposal 10.

Proposal 12 - Standards applicability

As mentioned in our comments relating to Proposal 2, we recommend that the AHRC becomes a nominating organisation for the existing international standards that are in development to ensure that its objectives are met.

Proposal 13 - Government taskforce

As mentioned in Proposal 2 and Proposal 12, we recommend that the AHRC becomes a nominating organisation for the existing international standards that are in development, and also recommend that the Commission conduct a review of existing ICT standards to understand how they could enable the concept of "human rights by design" in the development of all new technological solutions, not just those with AI in them.

Proposal 14 - Impact assessment tool & Question E

We recommend that the AHRC becomes a nominating organisation for the existing international standards that are in development via Standards Australia and that the concept of an impact assessment tool be pursued within that forum.

A tool such as the one being proposed should be driven through an internationally recognised standardisation effort and not specifically by an individual party in an individual country, considering the impact its introduction could have on the international competitiveness of our domestic economy.

Proposal 15 - Regulatory sandbox & Question F

CBA considers that engagement with stakeholders using forums such as a sandbox can be useful approaches for discussing and considering explainability and other issues associated with AI systems, including as they relate to machine learning and emerging explainability techniques.

Proposal 16 - Education within national strategy

We support this proposal to the extent the content has not already been developed within CSIRO's National AI Roadmap.

Proposal 19 - AI safety commissioner

We support the substance of this proposal. We recommend that the proposed role be assumed by the Australian Information Commissioner, considering the integral link between AI, data and privacy. We consider that another independent commissioner, with an independent focus and an independent remit would not be as effective in assisting academia, government or industry manage the challenges associated with using data safely and securely.

Proposal 21 - Inquiry into accessibility standard compliance

We support this proposal.

Proposal 22 - Government amendment to Broadcasting Services Act

We support this proposal.

Proposal 23 - Standard for accessible information

We support this proposal and recommend that the AHRC review existing standards for potential applicability.

Proposal 24 - NBN wholesale pricing proposal

We support this proposal. Organisations like CBA could also leverage its Benefits Finder feature in its CommBank app to assist with notifying customers.

Proposal 26 - Human rights by design education

We support this proposal if it is incorporated within internationally recognised standards.

Proposal 27 - Professional accreditation for human rights by design

We support this proposal if it is incorporated within internationally recognised standards.

Proposal 28 - Government commissioned organisation

We support this proposal.

Question H - Courses for human rights by design

In addition to engineering, science and technology, we propose that the AHRC include design and commerce/business management within the scope of any recommendation.

Proposal 29 - Digital communication technology standard

We support this proposal.

Conclusion

CBA supports the Commission's focus on the ethical development, deployment and use of AI and advocates for steps that encourage ongoing innovation and investment by Australian companies.

CBA welcomes the opportunity to have further ongoing discussions with the Commission regarding any aspect of this submission.