10 March 2020

Australian Human Rights Commission
Level 3
175 Pitt Street
SYDNEY NSW 2000

Lodgement: https://tech.humanrights.gov.au/

Dear Commission

Actuaries Institute Response to Human Rights and Technology Discussion Paper

The Actuaries Institute ("the Institute") is the professional body for Actuaries in Australia. The Institute is committed to promoting and maintaining a high standard of actuarial practice and contributing to public policy through policy submissions, thought leadership and expert analysis.

We thank the Australian Human Rights Commission for its extensive and thoughtful discussion paper on Human Rights and Technology (the "discussion paper"), and for the opportunity to give feedback on the contents of the paper.

Actuaries have a long and well-regarded history in considering the appropriate role of data and models in decision making, notably in the financial services industry, particularly insurance and superannuation. Many of our members have watched with interest the emerging global discussion around artificial intelligence (AI) ethics, and many have individually contributed to this debate. We consider that actuaries have an important role to play in such discussions, combining quantitative technical skills with commercial acumen, a professional code of conduct and a duty to serve the public interest.

Given the breadth of scope of the discussion paper, we only respond to aspects of it which we consider may create challenges for professionals or be of specific interest to a professional body such as the Institute. Since our points of feedback each relate to multiple questions asked by the discussion paper, we have not responded to questions individually, but identify in our response the various questions or proposals that our thematic points of feedback relate to.

The themes which we offer for your consideration are:

1. The unresolved conflicts between various forms of indirect discrimination, direct discrimination, and ideals of privacy.
2. The drawbacks of AI-specific regulation compared to generic legislation over decisions.
3. The appropriate comparative basis of AI and human decisions.
4. The role of professionals and expertise

Institute of Actuaries of Australia
ABN 69 000 423 656
Level 2, 50 Carrington Street, Sydney NSW Australia 2000
t +61 (0) 2 9239 6100 **f** +61 (0) 2 9239 6170
e actuaries@actuaries.asn.au **w** www.actuaries.asn.au

The unresolved conflicts between various forms of indirect discrimination, direct discrimination, and ideals of privacy

As a professional body, our Code of Conduct requires our members to comply with the various laws, regulation and professional standards which apply to their work. As such, we take particular interest in any accepted or perceived conflicts within the legislative regime.

Indirect discrimination within AI is something which has been extensively studied in recent years, generally under the banner of "fairness". Many recent examples of "AI bias" or "unfairness" are in fact examples of indirect discrimination. The discussion paper identifies several well-known examples (e.g. 6.3 (a) (i), p78; 6.5 (b), p85). The AHRC is correct in observing that it may often be difficult to detect indirect discrimination (6.5 (b), p84). We would go further and state that without access to data describing the protected attribute, it is generally not possible to demonstrate or, indeed, avoid indirect discrimination in a concrete manner. An analogous point is made on page 78 of the discussion paper, noting the various proposals for avoiding discrimination, which we observe typically require some use of data describing the protected attribute at hand.

Despite this need for data to avoid indirect discrimination, the prevailing environment discourages the collection of data describing protected attributes:

- With direct discrimination legislation also applying to each protected attribute, organisations typically do not wish to collect and therefore be seen to use the protected attribute for decision making, lest they be seen to be directly discriminating.
- With many protected attributes describing highly sensitive information about individuals and so effectively falling under the protection of privacy legislation, and with no obvious link for many such attributes to the decision procedure at hand, there may be no clear basis on which to collect such data.
- Legislation aside, socially it is likely to be unacceptable for an organisation to collect dozens of pieces of sensitive data relating to an individual in order to detect and avoid indirect discrimination against each individual protected attribute covered by the various State and Federal laws.

If data describing the protected attribute is not available, either for the reasons above or some other reason, it will not generally be possible for a practitioner to know whether indirect discrimination is occurring. Clearly this is of concern.

Furthermore, within indirect discrimination itself, there are now well recognised definitional conflicts which are not clearly recognised by the current legislation. Primarily, this revolves around incompatible definitions of harm. The discussion paper identifies an informative paper on the issue (Nayanan, 2018, as cited in 6.8 (b)(i), p90). We also note the various mathematical proofs and analysis of incompatibility of such definitions, for example Kleinberg et al (2017)[1] and subsequent work by the research community.

[1] https://arxiv.org/abs/1609.05807v2

With all this in mind, and noting the wide range of protected attributes covered by the legislation, the Actuaries Institute considers that under almost all decision making procedures today (AI or otherwise) people will be suffering from indirect discrimination to at least some degree, under some definition of harm, towards some protected class. Whether these existing procedures are allowable, noting some of the permitted defences under the discrimination acts for indirect discrimination, is often unclear. Many of our members are troubled by this, particularly in light of their professional obligations.

We suggest any review of the legislative regime (for example as outlined by Proposal 3) ought to contemplate these challenges and consider practical reforms. We suggest that one goal of the legislation ought to be that in the absence of specific case law relating to the situation at hand, a suitably trained professional should be able to clearly identify whether they are compliant with the law or not, and that seeking compliance with one aspect of the legislation should not come at the cost of non-compliance with the aims of other pieces of legislation.

The drawbacks of AI-specific regulation compared to generic legislation over decisions

The discussion paper makes a series of proposals reliant on a definition of "AI". As noted in the paper (e.g. 5.5 (a) (ii)) "AI" is not a well-defined term, hence the use of this within proposed legislation creates uncertainty around the boundaries of scope. Such uncertainty of scope will create inevitable arguments at law when an AI system is challenged under any such regime, and until case law or guidance develops, this will create uncertainty for professionals.

For the most part, we consider this to be avoidable. In many cases, the provisions proposed for AI systems could and perhaps should apply equally to all decisioning systems that meet the AHRC's proposed materiality threshold (having a legal or similarly significant impact on an individual). This would lead to a simpler, more powerful legislative regime, with less scope for debate and less uncertainty for professionals.

For example:

- Proposal 5 could simply require notification when such a decision is made *in general*, together with some relevant information about how the decision was made (for example whether it was automated or not), rather than making a notification rule specific to AI but no other mechanism of decision.
- Proposal 6 could apply to all decisioning processes, not just those that relate to AI.
- Proposal 7 could apply to all decisioning processes, with part (b) adapted to mean expertise in general around any complex mechanism of decision.
- Proposal 8, if adopted, seems reasonable to apply in any context, not just AI.
- If Question B is answered in the affirmative, this is perhaps also reasonable in general, not just for AI.
- Proposal 10, if adopted, seems reasonable to apply to all systems, not just AI systems.
- Proposal 13, notably the proposal for enforceable certification, if adopted should apply to all systems with the potential for harms material enough to warrant such a scheme, not just AI systems.

Generally, we can conceive of no clear example where rules ought to exist for an AI-informed decision but should not operate in some analogous fashion to an equivalent non-AI decision, and we suggest that if such cases do exist they would be rare exceptions which any legislation could specifically contemplate. To that end, we consider the definition promoted by Question A to be irrelevant for most legislative needs.

We note that the expansion could mean additional rules such as those proposed in the discussion paper would apply to many existing contexts which meet the materiality threshold – irrespective of whether AI is used in the decision or not. The potential impact of this, and any interaction with existing legislation, would require careful consideration. For example, decisions involving many existing financial products such as loans or insurance would potentially fall into this scope.

The potential for existing regulation of decisions to contain gaps created or exacerbated by "AI" appears to be part of the scope for the proposed Law Reform Commission work (Proposal 3), which we support. We suggest that conclusions about the relative merits of AI-specific or general legislation over decisions ought to wait for this review and form part of its scope.

The appropriate comparative basis of AI and human decisions

Most AI-informed decisions implemented today are not new decisions, but existing decisions to which AI is added as an enhancement or replacement. As observed in the discussion paper, AI may help us to better understand and to improve upon the biases inherent in many existing decisions but may carry the downside of creating new forms of bias (6.8, page 89). In our view, this illustrates an important opportunity: with AI-informed decision making we can precisely describe the outcomes we desire within software, which may be an improvement upon today. It also acknowledges that the current state of affairs is not perfect, and as we observed above, indirect discrimination of some form should be expected to exist within most existing decision procedures – AI or otherwise.

We caution against any regulation which assumes the status quo for a decision procedure is optimal for human rights outcomes. In our view, any such regulation may stifle the use of AI to enhance an existing decision context, leaving us with the existing process complete with its inherent biases and flaws. This is of concern to professionals who wish to incrementally improve upon today's outcomes using modern techniques.

With this in mind, we submit that:

- Proposal 6 should, where relevant, require an evaluation of the status quo as well as the proposed system, in order that it can be determined whether the AI system represents an improvement. It is possible that both the AI and human systems are found to be imperfect, but it may still be necessary to use one of them. In this situation, we suggest criterion (c) should not work only against the AI system but should apply with equal weight to the imperfect human decision, in order that incremental but still imperfect improvements are admissible.
- In response to Question D we suggest that human intervention does have the potential to detract from as well as improve outcomes. Hence any legislation of the form proposed should be addressed cautiously. For example, observing that an AI system creates some

form of harm does not necessarily justify a requirement for human intervention - which may make matters worse in some contexts.

Proposal 17, notably (b) and (c), again appears to assume AI can have negative or neutral impacts only on human rights. Where AI is implemented to enhance or substitute an existing decision, we suggest the evaluation process should be against the status quo, likely an imperfect situation. Evaluation should not be against some idealised state.

The role of professionals and expertise

Professionals and independent experts play an important role in bridging the information gap between lay parties. There are numerous examples of this today, with a range of professional bodies providing expert, independent advice to lay people on complex matters.

The actuarial profession has long been regarded, worldwide, as an important independent custodian of policyholder and broader public interest in the financial services industry, and has traditionally performed various certification roles involving material data driven decisions in this sector. For example, it is not practical for community stakeholders to interrogate the complex financial inner workings of an insurance company in order to assess the sufficiency of reserves, therefore actuaries are entrusted by regulators to balance the competing interests of various stakeholders and make independent recommendations for reserves using data and models, and governed by codes of professional conduct.

It may not always be possible, nor practical, for an organisation to make its decision processes completely transparent for audit by any interested party. This could be due to technical limitations, or a need to protect confidentiality and intellectual property in a competitive marketplace. In such cases, we believe there is an important assurance role for independent professionals, bound by a code of ethics, to play, to help bridge the information gap, and where necessary, to help to balance the competing interests of various stakeholders. We submit that actuaries may be able to play such a role in future.

With this in mind:

- In response to Question C, one mechanism to achieve "better access to technical information used in AI-informed decision-making systems such as algorithms" could be through an independent professional review of an algorithm.
- In response to Proposal 13, we are generally supportive of the concept of a certification scheme, however we caution against the following:
    - We are keenly aware of the sectoral expertise required for our certification activities today, and as such we do not support a "one-size-fits-all" approach to a design process or certification scheme. Sectoral expertise is essential to ensure that the nuances of the sector are considered in designing and evaluating a proposed system – what is reasonable in one domain is not necessarily reasonable in another. An example of a sectoral design consideration within human rights is the provision under the Age Discrimination Act for insurers to set prices and policy conditions based on the age of an insured, provided certain conditions are met. This acknowledges that age genuinely affects risk characteristics such as mortality, and so society might expect this as a reasonable exception to the Age Discrimination Act. A generic evaluation procedure of human rights concepts by a non-insurance expert may have failed to consider such matters. Many sectors are likely to have

similar nuances under any proposed scheme, which would require sectoral expertise to identify and evaluate.

  - Furthermore, we believe any formal requirement for a certification should be proportionate to the potential for harm of the algorithm. This will ensure scarce certification resources are placed where they can add most value, and algorithms with little or no potential for harmful outcomes or with comparably small scope of operation are not unduly hindered.

- Proposal 27 makes suggestions relating to continuing professional development. As a professional body we are generally supportive of this proposal and wish to highlight that the Institute is making available a range of CPD options in 2020 on discrimination for members. Additionally, our forthcoming major update to core modules of formal education will include an extended section on discrimination and human rights. We are keen to ensure that actuaries are appropriately knowledgeable and skilled in human rights issues, so as to ensure that as the global debate on AI ethics progresses, actuaries can continue to contribute to it.

We again thank the AHRC for an informative discussion paper and the opportunity to provide feedback. If you would like to further discuss this with us, please contact Elayne Grace, Chief Executive Officer of the Actuaries Institute, elayne.grace@actuaries.asn.au or on (02) 9239 6100.

Yours sincerely

Hoa Bui
President