

Consultation questions

1. What should be the main goals of government regulation in the area of artificial intelligence?

- 1.1 The question is based on the assumption that governmental regulation is needed in this sphere.¹
- 1.2 Most major firms developing or systematically using AI predictive analytics and other functionalities, and even start-ups, are developing internal corporate ‘fairness, accountability, transparency’ (FAT) policies and oversight processes. In this respect the ‘ethical AI’ or ‘responsible AI’ phenomenon resembles to some extent the wider experience of the corporate responsibility and sustainability one, with mostly voluntary, soft-law impulses and self-regulatory approaches, and associated concerns about these crowding out or pre-empting ‘harder’ ones. Like the wider corporate responsibility narrative, one concern relates to the materially different logics and consequences of an ethics-based approach as opposed to a legal (and human rights) based approach.² Other questions arise from the general experience of those working on corporate responsibility issues – relevant to the White Paper’s institutional design questions – about the relative strategic utility of the human rights vernacular.³
- 1.3 Where AI systems and platforms raise, as they do, human rights issues, reliance on industry self-regulation is clearly not sufficient.⁴ However, that is not to say that a regulatory strategy (and institutional design, the issue in this White Paper) should not seek deliberately to stimulate, condition and harness or enrol the ordering power of internal corporate systems in pursuit of the wider social objective.⁵ I return to this below.
- 1.4 A regulatory orientation (rather than a mere public policy or corporate ethics one) is required, and so some specialised institutional leadership is important, because of the impacts outlined in the White Paper.⁶ As Harari (2018) has written:

¹ That assumption is of course covered in the Commission’s July 2018 ‘Human Rights and Technology Issues Paper’ (and associated, still incomplete processes). Nevertheless, some discussion of wider issues is necessary here as context for the White Paper’s focus on a possible institutional structure for governing responsible AI in Australia.

² See further discussion at Q1.11 below.

³ See further discussion below.

⁴ See recently for example, and among others, Cath Corinne ‘Governing artificial intelligence: ethical, legal and technical opportunities and challenges’, (2018) 376 (2133) *Philosophy Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 1. Since algorithms are increasingly making decisions and so ‘regulating’ affairs including in public administration, the issue here is about the regulation of algorithmic regulation (Yeung 2018).

⁵ In the vein of various so-called ‘new governance’ approaches, including a responsive regulatory approach (e.g. Ian Ayres and John Braithwaite, *Responsive Regulation*, Oxford University Press, 1992). See Q1.9 below.

⁶ 2019, 7-8. See more generally the Commission’s July 2018 Issues Paper, Sections 3-4.

“... Since the corporations and entrepreneurs who lead the technological revolution naturally tend to sing the praises of their creations, it falls to sociologists, philosophers and historians ... [etc] to sound the alarm and explain all the ways that things can go terribly wrong.”⁷

In any event, some of the world’s largest and most influential technology and social media firms have called for regulation (rather than a corporate self-regulatory approach). Microsoft’s President Brad Smith recently told a high-profile AI summit: “We don’t want to see a commercial race to the bottom. Law is needed.”⁸

- 1.5 At the heart of human rights issues are questions of power in society, its distribution, abuse and non-accountability. Both the 2018 Issues Paper and the 2019 White Paper might have framed this issue by reference explicitly to the effect of AI on changing or entrenching forms of power (of the surveillance or judicial state but not necessarily of the regulatory state, and of private corporate power due to the greatly enhanced value of data and all-pervasive ‘surveillance capitalism’).⁹

In terms of Q1, a ‘main goal’ of regulation in this area is to use public regulatory power to check, in appropriate ways, how access to and use of AI technologies might radically transform power relations in society in potentially irreversible ways. This is so whether or not approached through the prism of human rights, which has its uses but also limitations. Future debate should arguably focus on this power dynamic issues rather than the particular framing of human rights impacts.

- 1.6 Harari has talked of the abdication of responsibility, in the sense that authority in society might come to be derived from an algorithmic source and only indirectly from a human source.¹⁰ Yet relevantly for this White Paper’s discussion about institutional forms in Australia on responsible AI there is another risk of abdication involved in the rhetoric of tech company CEOs about AI regulation. Salesforce’s Marc Benioff recently said:

“... the point of regulators and government [is] to come in and point True North”.¹¹

⁷ Yuval Noah Harari, *21 Lessons for the 21st Century* (Jonathan Cape, London, 2018), xiii.

⁸ Quoted in Cade Metz, ‘Is Ethical AI even possible?’ *New York Times*, 1 March 2019. See too the statements of CEOs of Uber, Salesforce, Facebook and Apple quoted in the White Paper at p. 5 (fn 4-7) that regulation is desirable, inevitable, needs to be more demanding, etc.

⁹ See too on this ‘power’ dynamic Martin Lodge and Andrew Mennicken, ‘Regulation of and by algorithm’ in Andrews et al 2017, 2; also the argument in Christiaan van Veen and Corinne Cath, ‘Artificial Intelligence: What’s Human Rights Got To Do With It?’ *Data & Society Points* (online), May 14 2018. Considerations of power differentials are central to Heung’s analysis (2018).

¹⁰ Harari 2018, 47; 57ff.

¹¹ Davos WEF 2018, quoted in the White Paper, 5 (n. 5).

It is hard to know if this is a serious statement. Notwithstanding the positives of this explicit corporate invitation to regulate and the merits of principles-based regulation setting broad value parameters,¹² this is a somewhat demoralising statement. It either reads like a form of abdication (we can't figure it out) or suggests a value-free corporate culture (we just don't know what values matter).

The problem with this CEO's statement is that 'True North' in AI is relatively easy to find without any governmental intervention. It relates to fundamental values such as privacy, non-discrimination and equal treatment, human dignity, universal human rights, accountability. It is worrying if corporate Australia also feels, like this CEO, a form of value paralysis absent some 'steer' or 'nudge' from government on basic societal values.

In this sense, the White Paper process should emphasise that whatever the outcome of debate about regulatory forms such as a Responsible Innovation Organisation, etc., it is hardly as if corporations building and using AI cannot, in the interim, discern and operationalise these basic values and legal parameters without top-down clearance from a central government.¹³

- 1.7 While more survey research may be needed, commercial actors are probably mostly concerned with whether a product or service using AI is perceived as trustworthy, so that it will be popular and successful. This is pro-social in general terms but is not the same as asking whether the AI platform (etc.) respects human rights or facilitates remedial measures. Absent a perceived legal liability framework, trust rather than conventional human rights impact is likely to be the factor driving corporate uptake of responsibility measures. If so, this difference in drivers or incentives for 'compliance' or behavioural change is something at the heart of any questions about the main goals of regulation and strategies to achieve those goals.
- 1.8 Without adopting a tone of panic or rushing to 'solutions', the White Paper perhaps might have made much more, in explicit terms, of the fact that time is of the essence. As a recent *Economist* article noted, regulators must respond to AI now: 2019 will require policymakers to start 'applying real thought to artificial intelligence'.¹⁴ Or as Harari has argued (2018):

"We cannot continue this debate indefinitely ... [v]ery soon someone will have to decide how to use this power [AI] – based on some implicit or explicit story about the meaning of life [and, in this context, of dignity in a human rights sense]... engineers are far less patient, and investors are the least patient of all.

¹² See further discussion below.

¹³ Of course, regulatees might require greater clarity and specificity about general principles, a familiar problem with relying only on a regulatory system that espouses these. See also remarks on mere lists of words that in fact have complex and contested meanings in any context, QX.X below.

¹⁴ Tom Standage, 'Regulating Artificial Intelligence' in 'The World in 2019' *The Economist* (December 2018), 22.

If you do not know what to do with the power [of these technologies, but also the power of how to govern them], market forces will not wait a thousand years for ... an answer. The invisible hand of the market will force upon you its own blind reply..."¹⁵

In this regard I set out below an argument for some form of governmental leadership organisation in this sphere to reflect the 'time-of-essence' factor in this fast-changing sphere, even if we have not yet resolved some of the very hard questions around conceptualising, designing and realising a suitable regulatory approach.

- 1.9 A regulatory orientation or presumption is needed, but what regulatory approach is to be adopted? This question is also the subject of the wider (July 2018 Issues Paper) Commission enquiry, but is directly relevant to the White Paper's discussion of one possible institutional form. A fulsome treatment is well beyond this Submission's scope (and is highly complex!).

However, one can venture a couple of points for consideration:

(a) Being comfortable with and embracing a messy pluralistic governance scene

As in Q1.3 above, any regulatory orientation must not just account for regulatory plurality as an empirical fact, but seek (as a normative proposition) to take advantage of this reality and the multiple entry-points that a rich, plural governance ecosystem provides for a smart and considered regulatory strategy. In this area as in so many others, internal corporate private rule systems (and adjudication systems e.g. on unpalatable content in social media), are part of the regulatory scene.¹⁶ So too are various private (or hybrid) and typically transnational normative systems capable of considerable influence on corporate behaviour, such as ISO standards.

Regulatory plurality (described variously by theories of polycentric, multi-level, networked nodal or plural governance) is not just inevitable but can be viewed as desirable. Policymaking is able to navigate the legitimacy concerns of explicitly harnessing private orderings in pursuit of public ends.¹⁷

What does this mean for a possible Responsible Innovation Organisation that may not have any regulatory mandate at all? It compels a mindset of making maximal use of existing orderings and systems to achieve policy goals.

¹⁵ Harari 2018, xiv. See too Matthew Risse, 'Human Rights & Artificial Intelligence: an urgently needed agenda', Harvard Kennedy School Working Paper, RWP18-015, May 2018.

¹⁶ See too the reference to corporations as 'enigmatic regulators' of their internal value systems, in the Issues Paper 2018, 23. These internal orderings are a challenge to governmental regulation, but also an opportunity.

¹⁷ Jolyon Ford, *Regulating Business for Peace* (Cambridge Univ. Press, 2015).

(b) Avoiding false dichotomies between different regulatory approaches

There is increasing scholarship (drawing analogies from existing regulatory theory) on the ideal type of regulatory approach or technique where ‘AI’ meets ‘Human Rights’. ‘Principles-based regulation’ has much to commend it, given the problems in this particular context with more detailed rule-systems.¹⁸ Often contrasted against this approach there is debate over a variety of ‘other approaches’¹⁹ (e.g. schemes that preference self-regulation and self-assessment, or co-regulation including with external audits and verification).

However, institutional design of the sort explored by this White Paper ought to avoid the premise that one must choose from among this smorgasbord of variously discredited regulatory theories or approaches. Too often one encounters arguments that some designs (principles-based regulation) would be universal but too general to offer proper guidance and certainty, while others (detailed rule-based orders) would quickly be outdated, gamed, etc.²⁰

What is missing perhaps is a sense of regulatory sequencing, and the co-existence (indeed co-dependence) of different regulatory approaches working together. The debate is cast in terms of alternative approaches, whereas many of these can co-exist, inform each other, precede each other. So a set of higher-order principles can sit behind and above any attempt, including at a later point in time, at more granular, contextualised (etc.) rule-making in particular sectors, contexts or in respect of particular concerns (e.g. privacy, or non-discrimination). Australia could start in the very near term with setting out some principles aligned with global efforts in this space (albeit arrived at through a process calculated to increase awareness and uptake of those principles). A Responsible Innovation Organisation, even one with an incomplete and incipient mandate, could begin to socialise these principles into industry and government even if it lacked (for now) any other regulatory complementarities.

This relates to the point above that time is of the essence. Yes, regulation in this area must be done well and got right, and must include powerful actors and voices

¹⁸ Principles-based regulation has its merits here (for instance, a regulatory system based on broad standards and values means that regulatees are forced to engage meaningfully in order to interpret these and act on them, and by doing so they [might] internalise values). Rule-based orders can be gamed, are too restrictive, or are quickly-outdated in fast-changing areas. See too Issues Paper 2018, 23, 32 (citing the work of Julia Black referred to in the Australian Law Reform Commission’s *Privacy Law* report (Report 108, 2009)). Of course, in setting any national-level general principles (the work of an ongoing Government project), Australia would want to align with high-profile joint efforts to establish ethical standards for AI. For instance, the IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems, established by IEEE Standards Association in 2017, and the Asilomar AI Principles (Future of Life Institute, 2017).

¹⁹ Issues Paper 2018, 24, 34.

²⁰ For a recent binary articulation of this, see Chowdhury ‘No one set of AI rules that will be granular enough to be applied [yet] high-level enough to be universal’ @ruchowdh, Twitter, 6 March 2019.

if it is to persuade and shape their actions, but it also cannot wait. With a sequential approach, it does not have to.²¹

As one development from the binary ‘principles-based regulation’ vs ‘rules-based systems’ debate, there is scope in the medium term for some meso-level standards that give more substance than general principles, but do not purport to be as prescriptive and detailed as rules. From a human rights-protective perspective, can we envisage and justify an intermediary or mezzanine level of principles-based non-rule standards for instance for the regulation of self-regulation (conditional self-regulation)? Moreover, is that something a Responsible Innovation Organisation – even one without yet any regulatory mandate, given a degree of urgency – could begin debate about generally and with particular sectors and policy spheres?

An international dimension (beyond the obvious cross-jurisdictional nature of governing AI)

- 1.10 In terms of the goal of regulation (Q1 here), there is of course an international dimension to all this. This is so quite apart from the fact that the nature of AI products and of the transnational firms developing and using these makes jurisdiction-based comprehensive regulation very complicated. The other international dimension is relevant to questions about a Responsible Innovation Organisation. It is that without a clear value-based and even regulatory posture of its own, in foreign policy terms it might be difficult for Australia to take coherent and principled positions on human rights and democratic governance issues arising in other countries where AI and other technologies are implicated (algorithm-assisted authoritative governance, and other ‘digital dictatorship’ scenarios).

A similar policy coherence issue would arise in relation to Australian defence procurement and offensive operational uses of technology abroad where human rights and humanitarian law issues arise.²²

Such considerations point towards the need in the short term, even if just provisionally, for some form of institution (whether or not a ‘Responsible Innovation Organisation’) -- or even a roving national and international Australian Ambassador -- capable of articulating values for and of Australians, including for reasons that might advance Australian firms’ competitiveness.²³

²¹ This also relates to the point made at Q1.6 above, debunking the Salesforce CEO’s suggestion that industry needs to wait until government shows ‘Due North’. That is, time is of the essence and Australia does need a set of signals and a storyline of sequential regulatory activities, but in other respects it is hardly as if industry is unable to discern and apply existing principles and frameworks to guide behaviour. Due North is not hard to find.

²² Meanwhile, Australia has an opportunity to show leadership in forums such as the UN Human Rights Commission on the AI + human rights agenda, something a Responsible Innovation Organisation could help to drive within the relevant parts of government.

²³ See ‘Economic reasons’ for an RIO in Q3 below (projecting a compelling and credible ‘brand Australia’).

Reinforcing the ethics vs human rights issue in relation to institutional responses

- 1.11 Whatever the interim or final form and structure of any institutional entity on Responsible Innovation, that body's establishment must reinforce the distinction between 'ethical AI' and 'responsible AI in a human rights compliance sense'.²⁴ This must be so even if 'ethical AI' is more topical and far less unwieldy. A number of problems exist with the 'ethics' paradigm in the AI context, given various human rights imperatives.²⁵ What is 'good' AI? What is 'fair'? What are the consequences of an ethics 'violation'?²⁶ Ethical concepts typically lack agreed definitions of the sort that law-enshrined human rights concepts do: use of AI, especially by public authorities in deciding administrative and judicial matters for individuals and groups, need to be based in the rule of law, not simply an ethical code.²⁷
- 1.12 Many leading companies' responses to the responsible AI challenge do not refer to human rights at all, let alone to frameworks such as the 2011 *UN Guiding Principles on Business and Human Rights*.²⁸ Any organisation envisaged in this space in Australia (along with the Commission itself, if they end up being different things) would need to bring that human rights dimension in explicitly, including by reference to concepts such as Human Rights Impact Assessments and Human Rights Due Diligence.²⁹ This is not to say that a human rights framing is *the* answer to addressing the underlying social problem or the regulatory goal of addressing it. As noted below, resort to human rights logics, institutions and discourse can have unintended consequences for strategies in this area. Corporations respond to certain incentives,³⁰ and it may well be that the 'ethical AI' paradigm (with its emphasis on self-governance, etc.) has considerable utility. Nevertheless, an Australian organisation would need to think hard before avoiding at least some explicit human rights dimensions to its regulatory design and its institutional messaging.
- 1.13 As noted in the 2018 Issues Paper, there is no challenge around articulating new human rights standards, but rather around applying these (including with a regulatory frame) to new tech challenges.³¹

²⁴ See too the Issues Paper 2018, 17 and fn 46: ethical approaches do not necessarily encompass a comprehensive approach from prevention to remedy (and lack the institutional infrastructure that the international human rights system has at national, regional and multilateral levels).

²⁵ Note 24 above. See too for example recently Christiaan van Veen and Corinne Cath, 'Artificial Intelligence: What's Human Rights Got To Do With It?' *Data & Society Points* (online), May 14 2018.

²⁶ Van Veen and Cath 2018, above.

²⁷ Phillip Alston, 'Statement on visit to the UK' (UN Special Rapporteur on Poverty and Human Rights), London, 16 November 2018.

²⁸ For one recent attempt to frame AI by reference to the UNGPs, see Filippo Raso et al, 'Artificial Intelligence & Human Rights: Opportunities & Risks' (Berkman Klein Center, Harvard, September 2018).

²⁹ Filippo Raso et al, *Artificial Intelligence & Human Rights: Opportunities & Risks* (Berkman Klein Center, Harvard, September 2018).

³⁰ See too White Paper 2019, 6. More work is needed on fleshing out what those incentives are, per sector, etc.

³¹ 2018, 11.

2. Considering how artificial intelligence is currently regulated and influenced in Australia:

(a) What existing bodies play an important role in this area?

(b) What are the gaps in the current regulatory system?

2.1 In terms of existing bodies, one could map all stakeholders. In the interests of (attempted...) brevity, a few comments suffice here.

Consumer ambivalence / unreliability

2.2 One body in society is AI users / consumers who, in some regulatory schemes around corporate responsibility, are themselves a potential regulatory resource. However, here it will be worth noting that consumers are in some respects conflicted or perhaps unreliable 'co-regulators': concerned about privacy, but wanting easy use and convenience. Scholarship on ethical consumerism in non-tech areas reveals this pattern fairly clearly.

Government leadership (beyond a Responsible Innovation Organisation)

2.3 Government is also a user and developer of these technologies, not just a prospective standard-setter and regulator. In an analogy from 'business and human rights' in the 'modern slavery in supply chains' context, federal and state governments can take a lead in setting certain conditions around the tender or contractual procurement of technology solutions, as can public providers of investment guarantees or project finance. This may shape market behaviour. One early role for an institution of the general sort envisaged in the White Paper might be to drive this agenda.

Gaps in regulatory systems

2.4 As discussed under Q1, it is clear that existing regulatory mechanisms and approaches are not commensurate with the task or challenge of algorithmic accountability or governance.³² As discussed in Q1.9 above, there are many potential insights from regulatory theory around the type of regulatory design best suited to AI (e.g. principles-based, design-based, outcome-based, risk-based, data-driven performance management, etc.).³³

³² Kevin Andrews et al, 'Algorithmic Regulation' (2017) CARR Discussion Paper no. 85 (CARR LSE / TELOS King's College London); Karen Yeung, (2017) 'Algorithmic regulation: a critical interrogation' (2018) 12(4) *Regulation & Governance* 505.

³³ For one overview, see Karen Yeung, 'Algorithmic regulation and intelligent enforcement' in Martin Lodge (ed.), 'Regulation scholarship in crisis?' (2016) CARR Discussion Paper no. 84, 50. See too Patrice Dutil and Julie Williams, 'Regulation Governance in the Digital Era: A New Research Agenda' (2017) 60 *Canadian Public Administration* 562; See too recently Allan Dafoe, 'AI Governance: A Research Agenda' Center for the Governance of AI, University of Oxford, August 2018.

- 2.5 One question is the adaptation of regulatory design ideas from other areas, which raises the initial question ‘in what ways is AI different from other contexts’ such that regulatory design analogies cannot be applied. The White Paper asks *what* it is we are looking to regulate or govern better.³⁴ Part of the challenge, which would affect the mandate of an regulatory or awareness-raising body that Australia arrives at, remains the problem with defining and covering ‘AI’ (let alone ‘new technology’) as a regulatory target.³⁵
- 2.6 A related issue from a human rights perspective is whether one is looking at impacts on civil and political rights, or more broad (but often related) impacts in a more distributional sense in terms of economic and livelihoods impacts.³⁶ Arguably, some of the distributional issues are best left to the democratic process rather than framed in AI regulatory terms.³⁷
- 2.7 One obvious regulatory gap is the lack of a dedicated body with a mandate on responsible technological innovation: see Q3 below. This is so even if that body has no regulatory mandate as such: as the White Paper adverts to, ‘regulation’ comprises any intentional activity to influence behaviour in pursuit of certain normative standards. A gap in our overall ‘regulatory’ system is thus a body that seeks to influence and inform behaviour in this space, irrespective of whether its final form makes it a federal regulatory agency with all the trappings of ASIC, the ACCC or other such bodies.
- 2.8 One particular regulatory gap noted in the White Paper relates to the review of adverse administrative or even judicial decisions made by or with algorithms. This is because it will not always be easy to discern the explanation for why an outcome was reached. The risk to a country with a long tradition of judicial review of administrative decision-making is that considered, case-by-case review opportunities might be replaced in some scenarios with the refrain ‘because that is what outcome the algorithm arrived at, so it must be right’. Ironically, technologies would then be creating greater distance between humans (and between the governors and governed) rather than facilitating the humanising aspects of an exchange based on seeking an explanation for outcomes. Of course these risks are manifold in authoritarian regimes.³⁸

³⁴ White Paper 2019, 5.

³⁵ Part of the challenge is the lack of information on the impact of AI/ML technologies, and of understanding by non-technical experts of the nature and impact of these: Lyria Bennett Moses, ‘How to Think about Law, Regulation and Technology: Problems with ‘Technology’ as a Regulatory Target’ (2013) 5 *Law, Innovation and Technology* 1. See too Filippo Raso et al, *Artificial Intelligence & Human Rights: Opportunities & Risks* (Berkman Klein Center, Harvard, September 2018) (difficulty of gauging impact, and proposing a framework for doing so).

³⁶ A ‘meta-question’ behind this is the utility of approaching the responsible AI governance question primarily as a human rights issue: see Q3 below.

³⁷ See in this regard Filippo Raso et al, ‘Artificial Intelligence & Human Rights: Opportunities & Risks’ (Berkman Klein Center, Harvard, September 2018).

³⁸ See 1.10 above (international dimension and ‘digital dictatorships’).

- 2.9 We live in an ‘age of transparency’ with high market and citizen/consumer expectation about responsiveness and openness by public and private entities. Yet algorithmic decision-making is not transparent, and since AI platforms can ‘learn’, even their programmers cannot predict or may not forensically understand the decision-making process. In what ways do regulatory design models need to change? The go-to model of the ‘new governance’ era (mandatory reporting regimes that rely on the market or consumers to police corporate behaviour) is not necessarily appropriate. Considerably more discussion and debate is needed on the appropriate mix of regulatory tools and techniques (along the continuum from education to persuasion to cooperation to coercion, and not forgetting remedy). However, resolution of this and of all the regulatory gaps in this area do not mean that it is too early to establish a body of the sort envisaged by the White Paper.
- 2.10 In terms of a general approach to regulatory gaps both in an interim and more ongoing sense, one area of possible focus is on the governance not of AI platforms (etc.) themselves, but of the underlying data-sets on which they rely, using relatively familiar paradigms such as data privacy regimes.³⁹ Huge data sets are needed to ‘train’ most AI systems. The data privacy agenda is thus key to the AI governance one. This suggests that rather than focusing on the direct regulation of responsible AI, greater emphasis be put on the enforcement and socialisation of data privacy rules.⁴⁰ A key challenge is that users / consumers do (or already have) consented to the use of their data with typically very little understanding about its potential onward use. We have some survey research on public perceptions of concern about data privacy and trust in tech / tech firms.⁴¹ However, this may be an area where policymakers keen to prevent a national-level crisis of confidence may need to lead even without an obvious groundswell of public opinion calling for greater enforcement of data privacy rules and the market for harvested data.
- 2.11 In terms of the regulatory ‘gap’, while Australia lacks and seeks to develop a national-level set of principles around responsible (ethical?) AI, in some respect there is no gap in relation to such principles, at least at the general overarching level.⁴² We can anticipate a further proliferation of lists of principles.⁴³ These

³⁹ Another way of looking at this is that future more legalistic regulation might seek to define principles, functions and requirements drawn from the experience (or anticipation) of using specific technologies, rather than provisions that attempt to regulate the specific technologies themselves: see Ivan Szekely and others, ‘Regulating the Future? Law, Ethics, and Emerging Technologies’ (2011) 9 *Journal of Information, Communication and Ethics in Society* 180, 183.

⁴⁰ See too Standage 2018, n. 2 above, 22.

⁴¹ See for example studies referred to in the White Paper 2019, 10 fn 28, and Issues Paper 2018, 15.

⁴² The point of an Australian set of general principles may be more about the value and impact of the process rolled out to arrive at this, rather than their particular content (balancing this need for ‘process legitimacy’ with the time-of-essence point made in this Submission).

⁴³ See very recently, for example, Cade Metz, ‘Seeking Ground Rules for AI’, *New York Times*, 1 March 2019, for a recent list of 10 recommendations for responsible AI by companies, simply listing qualities (Transparency; Disclosure; Privacy; Diversity; Bias; Trust [internal accountability and review systems]; Accountability [common

presuppose but will not necessarily resolve any regulatory gap. In part this is because they address the question of regulatory principles rather than tools, mechanisms and techniques to give these fairly easily agreed principles practical effect. An Australian institution of the sort envisaged in the White Paper can certainly play a role in developing, socialising and embedding Australian-made but globally-aligned principles for AI governance. However, it must avoid being party to a mere cascade of words such as ‘transparency’ and ‘accountability’ that everyone will agree are important, but the meaning of which (generally and in specific AI contexts) can be hugely contested.⁴⁴

3. Would there be significant economic and/or social value for Australia in a Responsible Innovation Organisation?

- 3.1 The White Paper explains clearly its decision to focus on a Responsible Innovation Organisation (whether, what form, etc).⁴⁵ Nevertheless, the Commission and others should remain alive to the risk that this initiative and focus might in some unfortunate way prejudice or artificially confine debate on the underlying questions, which remain ‘how and by whom are AI’s adverse social [human rights] impacts to be governed?’.
- 3.2 In responding to Q3, one view might be that since AI is so widely applicable in almost any field of life, one should not create a dedicated regulatory body but rather adapt existing rules (on privacy, discrimination, etc.) to ensure that they cover AI-related scenarios and implications.⁴⁶ However, this position confuses two different issues: (a) rule-systems (whether to have an AI-specific set of laws or regulations, or mainstream AI considerations into existing laws), and (b) institutions of governance (which may or may not administer some or all of these rule-systems). Even if Australia chooses not to develop a specific set of principles, standards or rules relating to AI (see Q1 above), it does not follow that there is no

set of standards]; Collective Governance [in self-regulation]; Regulation [companies work with government on developing standards]; Complementarity [tool for human use, not displacing humans].) One commentator has argued that the compilation of lists of general principles is part of a corporate instinct (‘tech ego’) to seek to scale and standardise everything (even regulation), but that we should stop being ‘self-congratulatory about thought leadership’ in this space, and start doing things (which are easier said than done): Chowdhury @ruchowdh, Twitter, 6 March 2019.

⁴⁴ This risk is present within human rights discourse too, including in this context: see Issues Paper 2018, 17 (‘P.A.N.E.L’ approach). It is a question of balancing the utility and traction of broad principles with acknowledgement of the risk that they remain a sort of symbolic, ritual rhetorical flourish: Jolyon Ford and Claire O’Brien, ‘Empty Rituals or Workable Models’ (2017) *UNSW Law Journal* 1223.

⁴⁵ The Foreword to the 2019 White Paper makes clear (p6) that is intentionally focusing on the single issue of a centralised organisation as one practical tangible contribution to the wider debate framed by the July 2018 Issues Paper.

⁴⁶ For example, Standage 2018, n. 2 above, 22, argues that it is a mistake, for this reason, to seek a dedicated AI governance / regulatory organisation such as the Food and Drug Administration in the US. In this regard see too Andrew Tutt, ‘An FDA for algorithms?’ (2016) 69 *Admin Law Rev* 83. The 2017 UK House of Lords report on AI governance observed that a comprehensive distinctive regulatory scheme for AI might not be viable or even necessary.

need for a dedicated organisation with a mandate relating to the governance of disruptive technologies.

- 3.3 The economic value of a Responsible Innovation Organisation (not just an 'Innovation Organisation') relates, in my view, to the significance of trust and reassurance in scaling up socially useful AI-based platforms and systems. Such a body may have a role to play in helping create a more level playing field, 'pre-competitive' space in the market for these things, and reduced barriers to entry for new actors in a field at risk of monopolistic behaviours. Such issues are beyond the scope of my expertise.

However, in global terms, a RIO might be useful to the idea of a values-based 'Brand Australia' that gives Australian innovators a competitive advantage in world markets for AI services because of the perception that Australian-regulated innovators have incorporated trust-generating, risk-mitigating ethical and human rights standards into their products and services.

- 3.4 In framing 'social value' we need also to think of the international / transnational context. While Australia ought not to look patronisingly at smaller Pacific neighbours, it does have some opportunity to help build democratic resilience and social cohesion in these states / societies by helping them to deal with the governance and other implications of new technologies such as AI. This could be one mandate for an RIO although is understandably likely to be of less priority than a national-level focus.
- 3.5 If the potential impacts of these technologies are as profound as the White Paper and Issues Paper suggest, the question is not really about social *value* as societal imperative. See Q5 below.
- 3.6 The proliferation of contexts across society in which these technologies do or might apply suggests that the Organisation's main role, at least initially while (enforcement-style and other) regulatory issues are debated, would be one of 'messaging' and the articulation of values including through coordination and fostering dialogue.

4. Under what circumstances would a Responsible Innovation Organisation add value to your organisation directly?

- 4.1 To the extent that the organisation led to or fleshed out some greater government articulation of regulatory design intent, it would add value by helping those designing a research agenda that might contribute to national-level imperatives and objectives in this field.
- 4.2 As a university researcher cooperating with Singapore and other countries in our region, having a clear federal government value-based position around AI in the research & innovation space would help to guide decision-making on collaborative

activities with overseas researchers based in political systems with potentially different value calibrations.

5. How should the business case for a Responsible Innovation Organisation be measured?

5.1 If we are seeing a *revolution* then does one really need to build a ‘business case’?

As observed in Q3.5 above, if the stakes are as high as the White Paper (and equivalent policy products elsewhere) suggests, in some respects a discussion framed in terms of a ‘business case’ seems rather narrow-minded or at least at odds with those stakes. ‘Business case’ connotes exploring the creation of a body or process with relatively everyday functions, where the only issues are routine budgetary and other trade-offs.

If the White Paper’s premises are correct, the implications for our society would suggest that we are not dealing with the sort of relatively routine bureaucratic line of enquiry involved in a making a ‘business case’. I say this entirely conscious of the stakeholder processes, fiscal politics and other practicalities of proposals such as this.⁴⁷

6. If Australia had a Responsible Innovation Organisation:

(a) What should be its overarching vision and core aims?

6.1 It is possible to sound too esoteric here, except when one appreciates the stakes. One influential input in the UK debate about organisational responses to responsible AI offered a profound overarching narrative (objective) around ‘human flourishing’.⁴⁸ There are good reasons to seek some degree of consistency with this well-expressed idea, even if it must be ‘home-grown’ and result from a localised process. At the end of the day, the questions that the White Paper is dealing with are about *how to live* and involve the collective imagination of a coherent, shared story or narrative, complete with certain values.⁴⁹

6.2 The questions in the White Paper about a RIO dwell on the nature of the organisation, whereas the key issue, if the underlying question is a regulatory one, remains the issue subject to the 2018 Issues Paper. This is about how should we

⁴⁷ The Commission may want to draw on studies of RRI (Responsible Research and Innovation) in addressing the White Paper’s questions. See for example Jack Stilgoe et al, ‘Developing a Framework for Responsible Innovation’ (2013) 42 *Research Policy* 1568; Bernd Stahl et al, ‘The Responsible Research and Innovation (RRI) Maturity Model: Linking Theory and Practice’ (2017) 9 *Sustainability* 1036.

⁴⁸ See White Paper 2019, 15, n. 43.

⁴⁹ This is the central theme of Harari 2018, quoted above.

design regulation: of what, at what points in time,⁵⁰ and by whom? The nature, structure and powers of a RIO are in most respects a secondary question to the prior question of what regulatory approach is needed. However, one role for an RIO in Australia – established in the very near term – might be to guide and indeed drive the ‘regulatory design options’ conversation (the prior question) since it is not obvious which other entity might do this other than, say, the Human Rights Commission which as a limited mandate, expertise and budget.

- 6.3 As noted in Q1, any such organisation must have, at the heart of its vision, an international or transnational orientation.⁵¹ While state uses (and abuses) of AI might be something for state-level governance, the most influential private firms in the world (and their products) are deeply transnational in operation and effect. It would make little sense for Australia to develop a regulatory or other body whose vision is confined to Australia (even if its jurisdiction is).
- 6.4 There are strong arguments for aligning the Australian approach to the UK one where feasible, or at least to a calculated perspective on the experiences of that society which is a few years ahead in deliberating these issues and establishing institutional entities such as are suggested in the White Paper.

The force and limits of the human rights paradigm / narrative

- 6.5 AI has many potential adverse implications for human rights. However, just because AI has human rights impacts does not mean the human rights framework and paradigm is the best (most effective, coherent, legitimate) vector for leading on socially responsible AI.
- 6.6 The White Paper originates in the Australian Human Rights Commission. One question that arises is whether a future RIO should have an explicit ‘human rights’ dimension, at least in terms of nomenclature and labelling. This is because while human rights issues arise at so many points in the AI phenomenon (as the White Paper amply shows), it is not obvious that the vocabulary, logics and mechanisms of human rights have, in Australia, the strategic transformative potential that one supposes. There may be unanticipated adverse consequences to relying too explicitly on a human rights basis for public and industry outreach on responsible

⁵⁰ That is, regulation of algorithmic decision-making in ‘real’ time or ‘reactive’ time? See for example Martin Lodge and Andrew Mennicken, ‘Regulation of and by algorithm’ in Andrews et al 2017, 4. Can we assume that reactive regulation will in time incentivise design corrections of real-time decisions as in any other area of administrative or judicial review?

⁵¹ See the White Paper, 9, acknowledging that responsible / rights-respecting AI is not solely an Australian challenge even if Australia can show some kind of lead.

/ ethical / rights-based AI.⁵² This is not a question of principle so much as a question of pragmatic issue advocacy-strategy.

(b) What powers and functions should it have?

- 6.7 In anticipation of regulatory activities (perhaps across a diverse range of bodies rather than a single regulatory one),⁵³ an RIO's main function in the near term may be largely symbolic. That is, it may be empowered to provide leadership across policymaking, business and investment, and civil society, around the responsible AI agenda: awareness-raising, best-practice policy or business guidance, expert insight (and reassurance or urgency/alarm as required), and so on.⁵⁴ The RIO could lead the debate on regulatory design that in some respects (as put in this submission) precedes the question of institutional design that the White Paper chose to focus on. This is consistent with the Issue Paper's conception of 'regulation' (as influence).⁵⁵
- 6.8 One would want to reflect, in terms of regulatory design conversations led by a RIO or ahead of its formation, on the lessons from the Banking Royal Commission 2019. This would be in terms of lessons on the regulation of very big and influential entities, as many tech companies are, as well as on the consequences of combining in the same agency regulatory powers (in an enforcement sense) with non-coercive ones (advice, research, awareness raising). This is true whatever the merits of various 'new governance' approaches that are not all about enforcement and punishment, since those approaches are likely to be apposite to AI governance.

(c) How should it be structured?

- 6.9 I have no particular expertise on this sub-question.
- 6.10 One observation may be that a structure with a clear single individual leading it, in the form of a Commissioner, may have more media and societal traction than an Organisation. The UK's experience under its *Modern Slavery Act* 2015 of having a high-profile proactive anti-slavery commissioner shows how this more

⁵² For a contrary view (human rights as a powerful framework and well-institutionalised vernacular in AI governance), see Christiaan van Veen and Corinne Cath, 'Artificial Intelligence: What's Human Rights Got To Do With It?' Data & Society Points (online), May 14 2018. For another recent argument on the benefits or imperatives of using the language (etc) of human rights in this sphere, see Mark Latonero, *Governing Artificial Intelligence: Upholding Human Rights & Dignity* (Data and Society, 2018). Yet even if public discourse on the social impact of technology is often implicitly or explicitly conducted in terms of human rights (2018 Issues Paper, 19) this does not necessarily answer the strategy question of how best to address the underlying risks and the place of human rights logics and frameworks in that regulatory and advocacy strategy.

⁵³ And in anticipation of the fallout and follow-up to initiatives such as the Australian national-level process on ethical AI, and as reflected in the 'Digital Platforms' Interim Inquiry Report, Australian Competition and Consumer Commission, December 2018.

⁵⁴ White Paper 2019, 8, 11 (and submissions to the July 2018 Issues Paper, referred to therein).

⁵⁵ See the 2018 Issues Paper, 22.

individualised office (supported by a secretariat, and in AI by a credible powerful Advisory Board) might have more impact on awareness and uptake than a more faceless organisation. Of course, the CEO of a Responsible Innovation Organisation might have the personality to take on this national roving commissioner role, but their ability to do so would be undermined by having to run the organisation itself. This suggests a leadership structure that separates the public leadership and messaging functions of such a body from the administrative functions. Different considerations would be required if the body had regulatory functions and associated enforcement powers, if any.

(d) What internal and external expertise should it have at its disposal?

- 6.11 A key issue here is the availability of expertise to help in design of regulatory options, let alone regularised or routine regulation. How is Australia to find the technical expertise at the nexus of law/regulation and algorithmic technologies? As with regulating markets, will regulators struggle to attract the best regulatory analyst talent in a fast-changing commercial innovation context?⁵⁶
- 6.12 How does one appoint a thought leader to head the RIO who is credible across business and policymaking, where the most expert knowledge is likely to be heavily engaged in the business of new technologies. This possible problem suggests a heavy role for an engaged Advisory Board that is drawn of experts who are unlikely to leave their current roles to head something like the RIO. The Human Rights Commission ought to be represented on that board.

(e) How should it interact with other bodies with similar responsibilities?

No response

(f) How should its activities be resourced? Would it be jointly funded by government and industry? How would its independence be secured?

No response

(g) How should it be evaluated and monitored? How should it report its activities?

No response

⁵⁶ See too Martin Lodge and Andrew Mennicken, 'Regulation of and by algorithm' in Andrews et al 2017, 4 on the need for a new kind of regulatory analyst.