WORK IN PROGRESS

ADM+S ARC Centre of Excellence for Automated Decision-Making and Society

ADM+S **WORKING PAPER SERIES**

What Role for Standards and Assurance in Regulating Artificial Intelligence in Australia?

December 2023

Authors:
Henry Fraser
Fiona Haines
José-Miguel Bello y Villarino
Kimberlee Weatherall
Christine Parker

WORKING PAPER 008

admscentre.org.au/publications

ADM +S

Acknowledgement of Country

In the spirit of reconciliation, we acknowledge the Traditional Custodians of country throughout Australia and their connections to land, sea and community. We pay our respect to their elders past and present and extend that respect to all Aboriginal and Torres Strait Islander peoples today.

Suggested citation

Fraser, H et al. (2023) What Role for Standards and Assurance in Regulating Artificial Intelligence in Australia?, ADM+S Working Paper Series 2023 (8), ARC Centre of Excellence for Automated Decision-Making and Society, DOI: 10.60836/n896-rt25

ARC Acknowledgement

This research was conducted by the ARC Centre of Excellence for Automated Decision-Making and Society (CE200100005), and funded by the Australian Government through the Australian Research Council.

Australian Government
Australian Research Council

TABLE OF CONTENTS

1. INTRODUCTION

The European Union's draft AI Act has captured the attention of technologists, policymakers and scholars of regulation and governance worldwide.[1] It aims to promote 'trustworthy' Artificial Intelligence ('AI') with regulations that apply to AI horizontally across industry sectors. It relies heavily on technical standards for the implementation of its main requirements. The text of the Act was initially proposed by the European Commission in April 2021, and both the Council of the European Union and the European Parliament have issued proposed amendments over the past year. The text is now under negotiation in a 'trilogue' between the Commission, the European Parliament and Council of the European Union, in accordance with European legislative procedure. [2] Along with a prohibition on certain intrusive or harmful uses of AI (such as use of general-purpose social scoring by government), the draft AI Act imposes a set of requirements for 'high-risk AI systems' that pose risks to safety or fundamental rights.[3] These include requirements of data governance, accuracy, human oversight, quality assurance, documentation and logging, explainability and risk management. Proposed amendments by the European Parliament set out requirements for general purpose AI systems and generative AI systems (such as the well-known ChatGPT chatbot), which would also be groundbreaking in the global regulation of AI.[4]

Following Europe's 'New Approach' to regulation (established in the late 1980s and cemented more recently in its 'New Legislative Framework') the Act provides that conformity with harmonised standards approved by European standardisation organisations creates a *presumption of conformity* with the Act's essential requirements.[5] Some AI providers may only need to conduct self-assessments, although the Act also provides for certification by

---

[1] *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, COM/2021/206 final, 21 April 2021. (Hereafter, Commission Proposal).

[2] For a side-by-side comparison of the three competing texts, see *Proposal for a Regulation of the European Union and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)* and *Amending Certain Union Legislative Acts* 2021/0106 (COD) Draft 20-0602023. Regarding the 'trilogue', see 'Trilogue', EUR-Lex Glossary (Web Page) <https://eur-lex.europa.eu/EN/legal-content/glossary/trilogue.html>.

[3] If the European Parliament's version is accepted, the definition of 'high-risk systems' will also include systems that pose risks to health and the environment.

[4] European Parliament, 'Parliament's Negotiating Position on the Artificial Intelligence Act' (June 2023) <https://www.europarl.europa.eu/RegData/etudes/ATAG/2023/747926/EPRS_ATA(2023)747926_EN.pdf>.

[5] See Art 40; Art 65(6)(b). See also Council of the European Union, 'Council Resolution of 7 May 1985 on a New Approach to Technical Harmonization and Standards' (7 May 1985); 'New Legislative Framework', *Internal Market, Industry, Entrepreneurship and SMEs* (Web Page) <https://single-market-economy.ec.europa.eu/single-market/goods/new-legislative-framework_en>; Lukasz Gorywoda, 'The New European Legislative Framework for the Marketing of Goods' (2009) 16 Columbia Journal of European Law. 161.

independent third parties for some AI systems.[6] Standards and certification processes will therefore play a significant role in determining how participants in the AI value chain manage risks from high-risk systems in Europe.

This paper considers the desirability of a European standards-based approach to the regulation of AI in Australia. The analysis is timely given the renewed interest in AI regulation signalled by the Department of Industry, Science and Resource's consultation on 'Responsible AI'.[7] One question in the consultation was whether governance measures taken by other countries are relevant, adaptable and desirable to Australia.[8] Another was whether changes to Australian conformity infrastructure might be required to support assurance processes.[9]

Australia has previously taken inspiration from European approaches to technology regulation. Australian product liability law and medical devices regulation both borrow heavily from Europe.[10] Recent proposals for reform to Australia's Privacy Act would also bring Australia more closely in line with Europe's General Data Protection Regulation.[11] The global influence of European privacy, data and technology regulation (sometimes described as the 'Brussels effect') is well known.[12] All of these factors, taken together with the detail, scope and ambition of Europe's approach to AI regulation, may encourage Australian policymakers to take inspiration from Europe once again.

This paper analyses the objectives of Europe's New Approach, assessing strengths, weaknesses, challenges and opportunities of relying on standards and assurance to manage risks from AI systems and to achieve socio-technical objectives such as 'responsible' and 'trustworthy' AI. It puts forward several reasons to avoid over-reliance on standards as a regulatory mechanism for AI. It is not clear that standards bodies have the right combination of

---

[6] See chapter 4, Commission Proposal.

[7] Department of Industry, Science and Resources, *Safe and Responsible AI in Australia* (Discussion Paper, June 2023) <https://storage.googleapis.com/converlens-au-industry/industry/p/prj2452c8e24d7a400c72429/public_assets/Safe-and-responsible-AI-in-Australia-discussion-paper.pdf>.

[8] Ibid question 5.

[9] Ibid question 13.

[10] David Harland, 'The Influence of European Law on Product Liability in Australia' (1995) 17 *Sydney Law Review* 336; Petahn McKenna, 'Australian Medical Device Regulations: An Overview' in *Handbook of Medical Device Regulatory Affairs in Asia* (Jenny Stanford Publishing, 2nd ed, 2018) ('Australian Medical Device Regulations').

[11] Attorney-General's Department, *Privacy Act Review Report* (Report, 16 February 2023) <https://www.ag.gov.au/rights-and-protections/publications/privacy-act-review-report, https://www.ag.gov.au/rights-and-protections/publications/privacy-act-review-report>.

[12] Anu Bradford, *The Brussels Effect: How the European Union Rules the World* (Oxford University Press, 2020) <https://books.google.co.uk/books?hl=en&lr=&id=mZXHDwAAQBAJ&oi=fnd&pg=PP1&dq=anu+bradford+brussels+effect&ots=DZ7lDO-SMJ&sig=WvW5ksthlLDMl9NhzA4_8B2Hi98> ('*The Brussels Effect*').

expertise and political legitimacy to credibly determine the value-laden questions of public policy that arise in the governance of AI systems that impact human rights and the environment. Nevertheless, this paper recognises that AI standards are developing apace and will inevitably play a role in AI governance in Australia. We therefore propose some principles for when and how standards are likely to make the best contribution to the regulation of AI. We contend that standards are most likely to be effective when the division of regulatory responsibilities between government, standards-makers and others matches their respective capabilities; when standards-making is as inclusive as possible; and when it appropriately reflects and accommodates uncertainty and lack of consensus about critical goals.[13] We also recommend prioritising concerted efforts to build the requisite capabilities across government, non-governmental organisations (such as standards bodies) and civil society.

2. THE NEW APPROACH

Europe's New Approach to technical harmonisation and standards, established in 1985, was designed to create a uniform approach to product safety.[14] It now applies to a wide range of products, from toys to boats to personal protective equipment.[15] The purpose of the New Approach was to create a harmonised approach to product safety throughout Europe, facilitating confidence and movement of goods in the European single market. Critically for our purposes, it was intended that this process would be the default means for new technological innovations and their risks to be assessed and managed.[16] The New Legislative Framework cemented the New Approach, while also establishing more detailed rules about its implementation – for example regarding the accreditation of notified bodies, and requirements for conformity assessment.[17]

---

[13] Rebecca Schmidt and Colin Scott, 'Regulatory Discretion: Structuring Power in the Era of Regulatory Capitalism' (2021) 41(3) *Legal Studies* 454 ('Regulatory Discretion').

[14] *Council Resolution of 7 May 1985 on a New Approach to Technical Harmonization and Standards — Publications Office of the EU* (n 5).

[15] *Council Resolution of 7 May 1985 on a New Approach to Technical Harmonization and Standards* (OJ C, C/136, 04.06.1985)< https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:31985Y0604(01)>; New Legislative Framework, Single Market for Standards — European Commission < https://single-market-economy.ec.europa.eu/single-market/goods/new-legislative-framework_en> accessed 28/9/23; See also Gorywoda (n 5).

[16] Ibid; 'New Legislative Framework' (n 5). See also Evaluation of the New Legislative Framework n — SWD(2022)364, European Commission <https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12654-Industrial-products-evaluation-of-the-new-legislative-framework_en>

[17] Gorywoda (n 5).

One of the hallmarks of the New Approach (as updated in the New Legislative Framework) is the role of standards in establishing conformity with the law. Manufacturers or independent third-party certifiers, known as 'notified bodies', undertake conformity assessments against certain essential requirements set out in the law to gain market access in the European Union.[18] However, to spare manufacturers the difficulty and uncertainty of interpreting and assessing conformity against these requirements, standards are used to simplify conformity assessment. Compliance with harmonised standards (standards developed by European standardisation bodies and approved by the European Commission) is voluntary, but it creates a presumption of conformity with legislative requirements. There is therefore a strong incentive to use standards to meet regulatory obligations, rather than to try to interpret and meet the more broadly drafted general requirements themselves.

The AI Act follows the New Approach / New Legislative Framework. This means providers of certain 'high-risk' AI and certain other parties in the AI value chain can meet their legal obligations by conducting, and in some cases even self-certifying, conformity assessments against harmonised European standards.[19] By contrast, other regimes for risky products such as pharmaceuticals require the approval of a regulator before a product may be placed on the market.[20] The European Commission has already issued a draft standardisation request to the European standardisation organisations CEN (European Committee for Standardisation) and CENELEC (European Committee for Electrotechnical Standardisation).[21] It has asked for standardisation deliverables (which may include technical reports and other explanatory documents as well as standards) on risk management, data governance, record keeping, transparency and information provision, human oversight of AI systems, accuracy specifications for AI systems, quality management and post-market monitoring, and conformity assessment. JTC 21, a joint technical committee of CEN and CENELEC, is responsible for the development of these standards on AI.[22] At this stage, it appears that many, though not all, of these standards will be adopted from international standards that have already been made available, or are

---

[18] *Council Resolution of 7 May 1985 on a New Approach to Technical Harmonization and Standards — Publications Office of the EU* (n 5) Annex II.

[19] Art 40; Art 65(6)(b).

[20] Michael Veale and Frederik Zuiderveen Borgesius, 'Demystifying the Draft EU Artificial Intelligence Act: Analysing the Good, the Bad, and the Unclear Elements of the Proposed Approach' (2021) 22(4) *Computer Law Review International* 97.

[21] *Draft Standardisation Request to the European Committee for Standardisation (CEN) and the European Committee for Electrotechnical Standardisation (CENELEC) in Support of Safe and Trustworthy Artificial Intelligence* (European Commission) <https://artificialintelligenceact.eu/wp-content/uploads/2022/12/AIA-%E2%80%93-COM-%E2%80%93-Draft-Standardisation-Request-5-December-2022.pdf>.

22 *Business Plan* for JTC 21 (CEN / CENELEC, 16 March 2022) <https://standards.cencenelec.eu/BPCEN/2916257.pdf>.

under development by JTC 1–SC 42, a joint technical committee of ISO (International Standards Organisation) and IEC (International Electrotechnical Commission).[23] Engagement by CEN–CENELEC in the ISO–IEC development process, followed by adoption of ISO–IEC instruments as European harmonised standard or the development of new standards, is consistent with CEN and CENELEC's ordinary practice, formalised in agreements with ISO and IEC.[24]

Part of the appeal of the New Approach is that it is supposed to allocate regulatory responsibilities according to capabilities. This kind of division is critical for effective co-regulation, where regulatory responsibilities are divided between private bodies (in this case, standards bodies and certifiers) and government.[25] European regulators, expert in public policy, give effect to policy objectives with technology-neutral 'essential requirements' in regulations or directives. Technically capable standards bodies determine technical implementation through the development or adoption of standards. And manufacturers (or in the case of the AI Act, AI providers) who best understand the conditions on the ground, take primary responsibility for conformity assessment.[26]

In practice however, the line between technical implementation and policymaking is always blurred.[27] This blurring of responsibilities is particularly pronounced in the AI Act. 'High-risk AI systems' are not like toys or boats or even medical devices. High-risk systems include systems used in law enforcement, university admissions, job recruitment and in determining access to essential services. They may also include systems posing environmental risks. Systems of this kind are socio-technical. The risk–cost–benefit profile of high-risk AI systems is complex and polyvalent. Whether such systems are responsible or trustworthy depends not only on how

---

23 Ibid. JTC 1 / SC 42 has published 20 such deliverables, with more than 30 others still under development. See Standards by ISO/IEC JTC1/SC 42 <https://www.iso.org/committee/6794475/x/catalogue/p/0/u/1/w/0/d/0>.

24 Agreement on Technical Cooperation between ISO and CEN (Vienna Agreement) 1991 https://isotc.iso.org/livelink/livelink/fetch/2000/2122/3146825/4229629/4230450/4230458/01__Agreement_on_Technical_Cooperation_between_ISO_and_CEN_(Vienna_Agreement).pdf?nodeid=4230688&vernum=-2.;

25 Christine Parker, *The Open Corporation: Effective Self-Regulation and Democracy* (Cambridge University Press, 2002) <https://books.google.co.uk/books?hl=en&lr=&id=jspFnQ8oqNcC&oi=fnd&pg=PR6&dq=christine+parker+2002&ots=F7BM02Dzjx&sig=gOSWiKkR-JVV_vdVdazAXTswrVY> ('The Open Corporation').

26 *Council Resolution of 7 May 1985 on a New Approach to Technical Harmonization and Standards — Publications Office of the EU* (n 5), Annex 2.See also 'Enhancing the Implementation of the New Approach Directives, European Commission COM(2003) 240 final, 7/4/2003 https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2003:0240:FIN:en:PDF, part 1.3.

27 See e.g., José-Miguel Bello y Villarino, 'Global Standard-Setting for Artificial Intelligence: Para-regulating International Law for AI?' (2023) 41 Australian Year Book of International Law 157; Merijn Chamon, *The European Parliament and Delegated Legislation: An Institutional Balance Perspective* (Bloomsbury Publishing, 2022) ('*The European Parliament and Delegated Legislation*'); Hans-W Micklitz, *The Role of Standards in Future EU Digital Policy Legislation: A Consumer Perspective* (ANEC BEUC, July 2023) <https://www.beuc.eu/sites/default/files/publications/BEUC-X-2023-096_The_Role_of_Standards_in_Future_EU_Digital_Policy_Legislation.pdf>; Schmidt and Scott (n 13).

they are developed and deployed, but also on how they are used, by whom, to what end and in what context.[28] Compare the challenge of ensuring a toy is safe and trustworthy to the challenge of ensuring that a high-risk AI system such as the controversial Suspect Targeting Management Plan used by New South Wales Police is safe and trustworthy.[29] For toys, the exercise is essentially one of balancing manufacturing cost against physical safety (risks of strangulation, choking, etc.). For a high-risk AI system, considerations not only of safety and cost, but also of privacy, non-discrimination, freedom of expression, employment rights, the right to a fair trial, and personal liberty may all be in play simultaneously. Assessing whether the accuracy, explanation, data governance and risk assessment for a high-risk system is appropriate will often engage competing rights of different individuals and groups, as well as public interests. And yet it is for these thorny assessments that technical standards are supposed to be used.

Because the Act sets out its essential requirements at a high degree of generality, and because compliance with standards creates a presumption of conformity, standards bodies will have considerable discretion in determining matters that may have weighty policy import.[30] There are bounds to this discretion, but control over standardisation by the European government tends to be indirect and attenuated. The terms of the European Commission's draft standardisation request on AI are, if anything, even more general than the essential requirements in the Act. European Commission representatives participate in technical committees as observers, where they must naturally exert some influence: but only indirectly. At the point of adoption (or rejection) of a standard, the Commission is limited to considering whether the standards correspond with the standards request, but this generally involves an examination of matters of form rather than substance.[31] And having drafted the essential requirements of the AI Act at a high level of generality (on the assumption that standards will provide the detail and certainty that stakeholders need), the Commission would find itself in a difficult position should it wish to reject the adoption of an AI standard. After adoption, the Commission has the power to object to harmonised standards.[32] In doing so, it can signal its regulatory preferences. By their nature, however, objections are piecemeal rather than being vehicles for coherent, carefully expounded

---

[28] Roel IJ Dobbe, 'System Safety and Artificial Intelligence', ed Justin B Bullock et al (Oxford University Press, Forthcoming) 1584 <https://arxiv.org/abs/2202.09292>.

[29] Vicky Sentas and Camilla Pandolfini, *Policing Young People in NSW: A Study of the Suspect Targeting Management Plan | Public Interest Advocacy Centre* (Youth Justice Coalition NSW, 2017) <https://piac.asn.au/2017/10/25/policing-young-people-in-nsw-a-study-of-the-suspect-targeting-management-plan/>.

[30] Veale and Borgesius (n 20).

[31] *Regulation 1025/2012 on European Standardisation 2012* Art 10 (5), (6).

[32] Ibid Article 11.

policy. The Commission cannot use observers, adoption or objections to supplant the standards process.[33] Once the line between essential requirements and technical implementation is drawn and matters with significant public policy implications are designated as matters of technical implementation, the lion's share of control over those matters lies with standards bodies.

3. ADAPTABILITY AND RELEVANCE OF THE EUROPEAN APPROACH TO THE AUSTRALIAN CONTEXT

The Department of Industry, Science and Resources has indicated an appetite for using standards, alongside risk assessments, as policy tools in a wide range of settings.[34] However, in its guidance to policy officers on using standards, it emphasises the need for policy officers to consider the suitability of standards for the policy setting and regulatory task at hand and to seek evidence of that suitability.[35] Whether standards are suitable for the task of regulating AI systems, especially high-risk AI systems with significant impacts on rights, is at best unclear. There are significant differences in the protection of human and consumer rights in Australia and Europe, as well as institutional differences. There are questions about the expertise and legitimacy of standards bodies to deal with certain aspects of AI governance, challenges for inclusiveness in standards-making and a lack of consensus about key goals in AI regulation. Finally, there are risks of misaligned incentives for certifiers and inadequate development of ecological standards. This part considers each of these impediments to emulating the European approach to AI and standards in Australia in turn.

Rights and Institutions

A regulatory framework must be considered in light of the pre-existing legislation and regulations that are found within any economic, cultural and political context. Transplanting legislation and regulations without doing so brings with it the very real risk that such transplants become 'legal irritants', causing counterproductive outcomes.[36] Europe has explicit, formal

---

[33] *The European System of Harmonised Standards: Q&A Commissioned by the German Federal Ministry for Economic Affairs and Energy* (Redeker Sellner Dahs, August 2020).

[34] *Best Practice Guide to Using Standards and Risk Assessments in Policy and Regulation* (Department of Industry, Innovation and Science, July 2016) <https://www.industry.gov.au/sites/default/files/June%202018/document/extra/best-practice-guide-to-using-standards-and-risk-assessments-in-policy-and-regulation.pdf>.

[35] Ibid 4, 9.

[36] Toby S Goldbach, 'Why Legal Transplants?' (2019) 15(1) *Annual Review of Law and Social Science* 583.

protection of human rights; for instance, in the Charter of Fundamental Rights of the European Union.[37] The whole AI Act proceeds from the assumption that fundamental rights are conceptualised and protected in that way. Risk to fundamental rights is one of the main things that makes a system 'high-risk', and thus required to conform with essential requirements. Australia has no charter of rights, far more limited and fragmented recognition of rights, less developed laws on privacy and data, and less exacting consumer laws. Even if Australia adopted precisely the same standards for AI as Europe, the lack of underlying regulatory infrastructures and capacities for protecting human rights and consumer protection would weaken the regulatory impact of AI standards. Whatever its strengths and weaknesses, Europe's New Approach has been in place for nearly 40 years, with concomitant development of institutions, expertise and processes for facilitating cooperation throughout the regulatory, standardisation and assurance ecosystems. While there is communication and cooperation between policy and technical organisations in Australia, this cooperation it is not formalised and regulated with the same degree of institutional support as Europe's New Approach.[38] In the case of AI, the potential for a transplanted European approach to become a legal irritant is high; in particular, that conformity with AI standards regulation might provide false reassurance: conferring unwarranted consumer confidence but without preventing significant failure.

Expertise and Legitimacy

In the context of high-risk AI systems, the technical implementation of essential requirements — such as requirements to implement risk management, establish appropriate human oversight or provide appropriate explanations of automated decisions — may have social, legal and political implications that are both significant and complex. It is not clear that standards bodies have the expertise required to take responsibility for matters that engage difficult and involved questions of politics, public policy, law and ethics.[39] It is telling, for instance, that the European Commission's Joint Research Centre has reported that the AI risk management standard developed by ISO–IEC (ISO–IEC 23894) does not adequately address risks to fundamental rights, health or safety for risk management. This has prompted CEN–CENELEC's JTC 21 to

[37] Charter of Fundamental Rights of the European Union (2009)

[38] Regarding cooperation in Australia's assurance infrastructure, see *Thematic Session on Standards: Moderator's Report* (World Trade Organisation — Committee on Technical Barriers to Trade, 5 December 2019).

[39] [REF Yeung]. Michael Veale and Frederik Zuiderveen Borgesius, 'Demystifying the Draft EU Artificial Intelligence Act—Analysing the Good, the Bad, and the Unclear Elements of the Proposed Approach' (2021) 22 Computer Law Review International 97.

begin its own work on an ambitious 'Checklist for AI Risk Management' under considerable time pressure.[40]

Standards bodies and other participants in assurance infrastructure (like accreditors and certifiers) generally have a technical, quantitative orientation, with engineers playing a prominent role. Socio-technical value judgments, however, require experience with public policy and may also call for expertise in a range of fields related to the humanities and social sciences. Experience with product safety, for instance, does not necessarily equip standards bodies to make decisions about how accurate a criminal recidivism predictor must be; when risks of racial or other bias have been sufficiently mitigated; what kind of documentation ought to accompany such a system; what arrangements for human oversight are most appropriate given the grave implications of predictions; or how best to explain predictions to different stakeholders such as affected individuals, courts and regulators.

Standards Australia has made commendable efforts to recruit people from a wide range of backgrounds, and with a wide range of expertise to participate in AI standardisation committees. In its Artificial Intelligence Standards Roadmap, for example, it recommended increasing the membership of the Artificial Intelligence Standards Mirror Committee in Australia to include participation from more sectors of the economy and society. Currently the membership of that committee is impressively diverse, to Standards Australia's credit, including academics, lawyers, policy experts and members of civil society organistaions. Without dedicated and sustained funding and support for this kind of participation, however, it is unclear how long this mix of expertise can be maintained.

Issues of political legitimacy also arise.[41] Notwithstanding questions of expertise, is it appropriate for lawmakers to delegate wide discretion about delicate questions of public policy to technocratic organisations with no democratic accountability?[42] This degree of discretion seems inconsistent with the division of responsibilities envisioned in the New Approach. It

---

[40] Josep Soler Garrido et al, *Analysis of the Preliminary AI Standardisation Work Plan in Support of the AI Act* (Joint Research Centre (European Commission)).

[41] Rotem Medzini and Karen Yeung, 'Background Paper: Assurance Regimes for Data-Informed Services' (22 May 2022); Karen Yeung, 'Operationalising Trustworthy AI Governance: Beyond Motherhood and Apple Pie?' (ADM+S Symposium, 20 July 2022); Henry L Fraser and Jose-Miguel Bello y Villarino, 'Acceptable Risks in Europe's Proposed AI Act: Reasonableness and Other Principles for Deciding How Much Risk Management Is Enough' [2023] *EJRR* ('Acceptable Risks in Europe's Proposed AI Act').

[42] Schmidt and Scott (n 13).Michael Veale and Frederik Zuiderveen Borgesius, 'Demystifying the Draft EU Artificial Intelligence Act—Analysing the Good, the Bad, and the Unclear Elements of the Proposed Approach' (2021) 22 Computer Law Review International 97.

seems very much to impinge into the domain of formulating essential regulatory requirements rather than purely technical matters of implementation.

Of course, Australia is not bound by the constraints of the New Approach but the matter of legitimacy of a co-regulation regime relying on standards still looms large. Standards tend to work best, and carry the greatest imprimatur of legitimacy, when there is consensus about goals and at least some degree of agreement and clarity about practical implementation.[43] Such consensus has still not crystallised for AI governance. Indeed, one of the most distinctive features of AI governance and policy is the persistently wide range of views both about the broadest questions of public policy (what is good and bad AI?) and the narrower questions of implementation (e.g., what is a good explanation of an AI decision? How should a principle of fairness be implemented?)[44] Not least because of this lack of consensus, there needs to be space for the airing of uncertainty and disagreement. Inclusion of and consultation with stakeholders are critical conditions for legitimacy in the exercise of regulatory discretion of the kind contemplated for standards-makers.[45] Members of the public and affected persons must at least be given an adequate opportunity to comment on the regulatory instruments that will affect them (in this case, standards) and ideally should be involved throughout the whole lifecycle of those instruments, from their inception to their regular review.[46]

Inclusiveness and Access

Unfortunately, neither the creation nor the use of standards is very inclusive.[47] The Regulation on European Standardisation, the Commission's draft standardisation request on AI and the European Parliament's proposed amendments to the AI Act all emphasise the need to promote stakeholder participation in standards-making, including participation by civil society organisations. [48] Standards Australia has also, as mentioned, invited civil society

---

[43] Julia Black, *Rules and Regulators* (Clarendon Press 1997)

[44] Kristin Undheim, Truls Erikson and Bram Timmermans, 'True Uncertainty and Ethical AI: Regulatory Sandboxes as a Policy Tool for Moral Imagination' (2023) 3(3) *AI and Ethics* 997 ('True Uncertainty and Ethical AI'); Maria Nordström, 'AI under Great Uncertainty: Implications and Decision Strategies for Public Policy' (2022) 37(4) *AI & SOCIETY* 1703 ('AI under Great Uncertainty'); Lindsay Weinberg, 'Rethinking Fairness: An Interdisciplinary Survey of Critiques of Hegemonic ML Fairness Approaches' (2022) 74 *Journal of Artificial Intelligence Research* 75 ('Rethinking Fairness').

[45] Schmidt and Scott (n 13) 467.

[46] Schmidt and Scott (n 13); Karen Lee, 'Legitimacy in the New Regulatory State'.

[47] Christine Galvagna, *Discussion Paper: Inclusive AI Governance* (Ada Lovelace Institute, 30 March 2023) <https://www.adalovelaceinstitute.org/report/inclusive-ai-governance/> ('*Discussion Paper*').

[48] *Regulation 1025/2012 on European Standardisation* (n 31). European Commission (2022) *AI Act: Draft Standardisation Request*, Annex II, accessed 14/4/23 <https://artificialintelligenceact.eu/wp-content/uploads/2022/12/AIA-%E2%80%93-COM-%E2%80%93-Draft-Standardisation-Request-5-December-2022.pdf>; *Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the*

representatives to participate on its AI standards committees. However, barriers to participation in the development of standards tend to be practical rather than formal. Joining a committee to work on a given standard (e.g., at ISO or Standards Australia) is generally straightforward. The problem is that large companies have the resources to support consistent, ongoing participation and networking by their representatives, while stakeholders such as civil society organisations and small- and medium-sized businesses do not.[49] As a consequence, large commercial interests tend to have a disproportionate influence over the development and content of standards.

There is also a downstream inclusiveness problem: a problem of cost and ultimately of access to law, with implications for the rule of law. Whereas it is free to access legislation (though challenging for ordinary citizens to interpret the law), access to standards comes with a price tag. Standards Australia charges approximately AUD814 per year for package subscriptions to standards in particular fields such as construction.[50] Individual ISO–IEC standards deliverables such as technical standards and technical reports cost upwards of 100CHF (AUD 173), with the technical report on bias in AI systems, for example, currently priced at 166CHF (AUD294).[51] Paying this kind of money for even a fraction of the more than 50 ISO–IEC standards on AI that are published or under development would be prohibitive for many stakeholders.[52] And yet, the least well-resourced AI developers are precisely the group that the standards-based presumption of regulatory conformity is supposed to help. A basic tenet of the rule of law is that laws must be accessible, and the process by which they are made must be open and transparent.[53] An arrangement where the presumption of conformity with the law depends on adherence to standards that are made behind closed doors, and whose contents stakeholders are not even allowed to know without first paying substantial sums of money, would sit ill with the rule of law.

European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts s P9_TA(2023)0236 <https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html>, Amendments 103 and 104

[49] Galvagna (n 47) 42. See also Hans-W Micklitz, 'The Role of Standards in Future EU Digital Policy Legislation: A Consumer Perspective' (ANEC BEUC 2023) <https://www.beuc.eu/sites/default/files/publications/BEUC-X-2023-096_The_Role_of_Standards_in_Future_EU_Digital_Policy_Legislation.pdf> accessed 26 July 2023; Henk J de Vries et al, *SME Access to European Standardization Enabling Small and Medium-Sized Enterprises to Achieve Greater Benefit from Standards and from Involvement in Standardization* (2009)..

[50] Small Business Sets, Standards Australia https://store.standards.org.au/sets

[51] See 'Standards by ISO/IEC JTC 1/SC 42', ISO <https://www.iso.org/committee/6794475/x/catalogue/p/1/u/1/w/0/d/0>

[52] On cost as a barrier to access to standards, see e.g., de Vries et al (n 49).

[53] *Rule of Law* (Parliamentary Education Office) <https://peo.gov.au/understand-our-parliament/how-parliament-works/system-of-government/rule-of-law/>.

Risks of Misaligned Incentives

As with all forms of regulation and governance, standards and assurance processes are susceptible to capture and misaligned incentives.[54] Industry participants in standards-making have an interest in including their own practices and preferences into standards, lowering their own costs and raising the costs of their competitors.[55] Self-certifiers clearly have conflicts between their commercial interests and the public interests supposed to be protected through the certification process. Third-party certifiers also deal with potentially conflicting interests and incentives. Certifiers often owe duties of confidentiality to their clients, meaning they are not able to disclose risks that they detect but which are outside the remit of their certification.[56] Certifiers do not necessarily certify that a product, site or service is safe in substance — in some cases they are required only to certify that the assurance process has been followed in form.[57] In the worst case, narrow, formalistic standards and certification processes paper over risks and bad practice — as occurred in the notorious Rana Plaza and Kader Toy Factory disasters.[58]

Sustainability Gap

Environmental impacts have been something of an afterthought in 'responsible AI' discourse and governance. Yet, the extent of environmental damage potentially caused by AI systems is striking. Key impacts include the energy and water use of the data centres in which data is stored and machine learning models are trained and applied, the impact of mining and production of graphical processing units and other equipment needed to develop and run AI systems, and the waste created by regular replacement of equipment.[59] References to

---

[54] See generally Daniel Carpenter and David A Moss, *Preventing Regulatory Capture: Special Interest Influence and How to Limit It* (Cambridge University Press, 2013) ('*Preventing Regulatory Capture*').

[55] Maurits Dolmans, 'Standards for Standards European Union Law' (2002) 26(1) *Fordham International Law Journal* 163, 171.

[56] We are indebted for this insight to participants at a workshop, convened under Chatham House rules, on 'Assurance Regimes for Data-Driven Services' at the University of Birmingham on 22 May 2023. The workshop was hosted by Professor Karen Yeung and Dr Rotem Medzini from the University of Birmingham as part of a project led by Yeung pursuant to the European Lighthouse on Secure and Safe AI Network of Excellence www.elsa-ai.eu

[57] See e.g., *Ku-ring-gai Council v Chan* [2017] NSWCA 226.

[58] See, e.g., 'More for show than safety: Certificates in the textile industry', European Centre for Constitutional and Human Rights <https://www.ecchr.eu/en/case/more-for-show-than-safety-certificates-in-the-textile-industry/>

[59] E.g., Sasha Luccioni, 'The mounting human and environmental costs of generative AI', Ars Technica, 4/12/2023. Available at: https://arstechnica.com/gadgets/2023/04/generative-ai-is-cool-but-lets-not-forget-its-human-and-environmental-costs/; Ligozat, Anne-Laure, Julien Lefevre, Aurélie Bugeau & Jacques Combaz. 2022. 'Unraveling the Hidden Environmental Impacts of AI Solutions for Environment Life Cycle Assessment of AI Solutions.' *Sustainability* 14, no. 9: 5172. https://doi.org/10.3390/su14095172

environmental risk were a late inclusion in the European Parliament's proposed amendments to the EU AI Act. After representations by NGOs and Green Party members of the European Parliament,[60] the European Parliament's June draft of the Act now recognises being 'environmentally friendly' as a priority of AI regulation[61] and also includes provisions that would require providers of AI systems deemed to be high risk to produce an environmental risk assessment (Article 9.2a, p55) and to make use of appropriate standards to reduce environmental impact, particularly energy use, resources and waste (Article 28b.2(d), pp39–40). This would require making significant new standards to quantify and log energy use and other environmental impacts of AI systems.[62] However, it is striking that the environmental impact of any end use of AI is almost entirely out of scope. For example, an environmentally friendly AI may be used to identify new fossil fuel mining sites or to program and personalise advertising for unnecessary consumption.[63]

4. WHEN STANDARDS ARE LIKELY TO WORK

Despite the foregoing concerns, it is clear that standards will play a role in the governance of AI. ISO–IEC has published 20 standards on AI already, with more than 30 currently under development.[64] IEEE has a broad portfolio, which includes publicly accessible foundational AI standards.[65] Standards Australia is involved in the development of the ISO–IEC standards and is likely to adopt them. More generally, for engineers and computer scientists, standards are the main interface with regulation and governance and indeed one of the main sources of design information.[66] Therefore, it is worth understanding when standards are likely to work best.

At the heart of the New Approach is a recognition that regulatory responsibilities and discretion should be allocated according to capabilities. Standards overwhelmingly deal with

---

[60] AlgorithmWatch. 2023. 'EU Artificial Intelligence Act -- Recommendations on Ecological Transparency,' https://algorithmwatch.org/en/wp-content/uploads/2023/05/2023-05-02-EU-Artificial-Intelligence-Act---recommendations-on-ecological-transparency.pdf

[61] European Parliament 2023a

[62] See OECD. 2022. *Measuring the Environmental Impacts of Artificial Intelligence Compute and Applications: The AI Footprint* https://www.oecd.org/publications/measuring-the-environmental-impacts-of-artificial-intelligence-compute-and-applications-7babf571-en.htm

[63] For an example seeking to address what AI is used for and its environmental impact, see Simon Coghlan and Christine Parker, Harm to Nonhuman Animals from AI: a Systematic Account and Framework. Philos. Technol. 36, 25 (2023). https://doi.org/10.1007/s13347-023-00627-6

[64] 'Standards by ISO/IEC JTC 1/SC 42', ISO < https://www.iso.org/committee/6794475/x/catalogue/p/0/u/1/w/0/d/0>

[65] 'AIS Standards', IEEE SA https://standards.ieee.org/initiatives/autonomous-intelligence-systems/standards/

[66] Bonnie Osif, 'Make It Safe and Legal' in *Integrating Information into the Engineering Design Process* (Purdue Univ. Press, 2014) 115 <https://library.oapen.org/bitstream/handle/20.500.12657/24969/1005134.pdf?sequence=1#page=126>.

specifications, procedures and guidelines aimed at promoting safety, consistency and reliability.[67] It is in relation to these qualities and arrangements — safety, consistency, reliability and process — that standards are likely to work best.[68] They are likely to be useful in developing and documenting good practice in relation to technical aspects of AI governance, including data governance, documentation and logging practices, algorithmic inspection and audit arrangements, training and testing, and establishing common metrics for accuracy and robustness. In these domains where technology develops rapidly, standards also have the advantage of being regularly updated.[69] Standards may also help in setting basic performance criteria for fairness, explainability and oversight, but perhaps to a lesser extent, as these are more open-ended goals.

More generally, standards, certification, audit and other assurance practices may be useful in providing assurance that appropriate processes have been followed in the development, deployment, and use of AI systems. For example, human rights impact assessments — which are likely to be standardised — may provide assurance that AI developers have at least considered human rights impacts in a systematic way, even if they do not necessarily ensure that human rights impacts are managed in the best possible way (especially since there will be disagreements about how to balance competing considerations). Standards and certification may also provide assurance that appropriate organisational measures are in place, with responsibilities for AI risks allocated to appropriately senior and qualified people within an organisation.[70] With any practice- or process-based standard however, questions may still remain about whether process and practices reliably produce the more substantive goal (such as trustworthiness or fairness) that is intended.[71]

Accountability is also likely to be important in avoiding some of the pitfalls of standards and certification. For instance, certifiers are less likely to treat certification as a formalistic box-checking exercise if they owe duties of care and could face potential liability in cases where

---

[67] See Jose-Miguel Bello y Villarino et al, *Standardisation, Trust and Democratic Principles: The Global Race to Regulate Artificial Intelligence* (United States Studies Centre, 31 July 2023) <https://www.ussc.edu.au/standardisation-trust-and-democratic-principles-the-global-race-to-regulate-artificial-intelligence> (‘*Standardisation, Trust and Democratic Principles*’).

[68] Fraser and Bello y Villarino (n 41).

[69] *Best Practice Guide to Using Standards and Risk Assessments in Policy and Regulation* (n 34) 6.

[70] See e.g., NIST, Artificial Intelligence Risk Management Framework (AI RMF 1.0) AI 100-1 (January 2023) <https://www.nist.gov/itl/ai-risk-management-framework>, p 8.

[71] Kira JM Matus and Michael Veale, ‘Certification Systems for Machine Learning: Lessons from Sustainability’ (2022) 16(1) *Regulation & Governance* 177 (‘Certification Systems for Machine Learning’).

certified products or services fail.[72] However, as things stand in Australia, it is not clear that such duties fall upon certifiers as a matter of course. For instance, the New South Wales Court of Appeal has held that a local council's building certification authority did not owe a duty of care to purchasers of a property to take reasonable care to prevent loss to the purchasers as a result of the defective work by the previous owner's builder. The role of a final occupation certificate, the court held, is to show suitability of a building for occupation and use and this 'does not require that all of the building work that is the subject of the development consent has been carried out in accordance with approved plans and specifications, and in a proper and workmanlike manner'.[73] In other words, where the certification exercise is framed narrowly — providing assurance only about very specific matters — certifiers may escape liability for certifying systems that fail to meet commonsense expectations of safety and trustworthiness.

5. PRIORITIES AND QUESTIONS

Australia does not have to follow the European approach to AI regulation, where conformity with essential regulatory requirements is determined by compliance with standards. It is yet to be seen whether standards are up to the task set for them by Europe's draft AI Act. AI governance, by whatever means, is inherently multidisciplinary and involves difficult socio-technical questions. There are serious doubts about whether technical standards can really be the basis for a judgment — indeed, a statement of regulatory conformity — to the effect that the whole complex system with all its inputs, impacts and human factors is safe, responsible and trustworthy.

Australian policy makers will need to answer several pressing questions over the coming years as our AI governance ecosystem develops. How comfortable are Australians with relying on technical standards bodies to answer questions of public interest about rights, the environment, discrimination, and so on? If not standards bodies, who has the legitimacy and expertise to set rules and policy in relation to AI risk acceptability, explanation and other aspects of AI governance with significant public policy implications? If this combination of expertise and legitimacy does not yet exist in Australia's regulatory ecosystem, this may be a reason to establish a dedicated AI regulator.[74] Further research and policy discussions are

[72] Jan de Bruyne, 'Certifying AI — some challenges and the role of liability', Standards and Certification for Trustworthy Data-Driven Technologies, ADM+S / University of Birmingham, 23 May 2023.

[73] Ku-ring-gai Council v Chan [2017] NSWCA 226

[74] See the recommendation to establish an AI Safety Commissioner in Australian Human Rights Commission, *Human Rights and Technology* (Final Report, 2021) part 10.

needed to develop a clearer sense of the role of standards bodies in highly charged policy decisions and to define the complementary elements needed in a regulatory ecosystem to ensure that socio-technical aspects of AI governance are performed with appropriate expertise and legitimacy. There is also an urgent need to build up the requisite expertise throughout the AI governance ecosystem, including among government agencies, standards bodies such as Standards Australia, accreditors and certifiers, and throughout the Australian economy.

Government may not be best placed to provide detailed guidance on *technical* aspects of AI (as recognised by Europe's New Approach). It is, however, better positioned than standards and assurance professionals to provide detailed guidance on how stakeholders should grapple with AI's *socio-technical* aspects, such as dealing with trade-offs between the rights and interests of different stakeholders. Indeed, some trade-offs may be incommensurable and there will be disagreements about the risks that a system poses. Effective governance measures would recognise and accommodate such complexities and tensions transparently, rather than seeking to promote trust through checkbox certifications that fail to truly reckon with whether a system is trustworthy in all the circumstances.[75] Again, it is government that is best placed to take responsibility for facilitating this kind of deliberative, open-ended public policy.[76] Even where standards are adopted or incorporated by reference into Australian law, the government may need to provide additional guidance to stakeholders on difficult public policy questions.[77] And where a given standard or part of a standard does not meet Australian legal or ethical requirements, the government can specifically issue guidance to this effect (as the European Commission does with technical standards).[78] If Australia establishes a dedicated AI regulator, or even an AI Safety Commissioner without enforcement powers (as recommended by the Human Rights Commissioner), guidance on difficult socio-technical questions involved in AI governance and commentary on technical standards and other forms self-regulation should be part of its remit.[79]

---

[75] Andy Stirling, 'Keep It Complex' (2010) 468(7327) *Nature* 1029; Fiona Haines, *The Paradox of Regulation: What Regulation Can Achieve and What It Cannot* (Edward Elgar Publishing, 2011) ('*The Paradox of Regulation*').

[76] Parker (n 25).

[77] See e.g., Henry L Fraser and Jose-Miguel Bello y Villarino, 'Where Residual Risks Reside: A Comparative Approach to Art 9(4) of the European Union's Proposed AI Regulation (Working Paper, September 30, 2021) <http://dx.doi.org/10.2139/ssrn.3960461>, part 5.

[78] For an account of one particular objection to a technical standard issued by the European Commission see Fraser and Bello y Villarino (n 41) part III.

[79] Australian Human Rights Commission (n 74).

Capacity building is also crucial. The government has hitherto emphasised the building of technical and economic capacity to take advantage of opportunities presented by AI.[80] The same capacity building is required to ensure effective governance. Government would be well advised to develop and implement guidance and training on fundamental rights, public health, environmental and other AI impacts for accreditation bodies, certifiers and others involved in AI governance. This should be in addition to training on technical aspects of safe and responsible AI such as accuracy, robustness and data governance. Among the constellation of Australian institutions, universities are unique repositories of multidisciplinary knowledge and expertise about cross-disciplinary collaboration. Partnerships between universities and regulators, standards-makers, accreditation bodies, certifiers, civil society and industry are likely to be valuable in such an endeavour.[81]

Representativeness and inclusiveness in standards-making and certification could be enhanced by various means, including through government funding to assist civil society and academic participation in standards-making and assurance and government funding of new, unconventional standards-making bodies.[82] A recent report by the Ada Lovelace Institute, for instance, recommended increasing the range of organisations eligible for mandated participation in standards-making in Europe and supporting their participation, as well as the participation of other civil society organisations with dedicated funding.[83] There is value in a diverse standards and assurance ecosystem with competing standards and assurance regimes. Australia should continue to invest both in the development of local standards and in fostering Australian participation in global standards-setting for AI. Indeed, as standards-making for AI is currently in an early stage, there is an opportunity to exercise influence by setting aside resources to support Australian AI experts' participation in standards-making.

6. CONCLUSION

While there is much to admire in Europe's commitment to trustworthy AI, and much to learn from the various processes and institutional arrangements that the AI Act envisions, Australia should be cautious about adopting the European approach to standards for AI. As this paper

---

[80] *Australia's Artificial Intelligence Action Plan* (Department of Industry, Science and Technology, 16 August 2022) <https://webarchive.nla.gov.au/awa/20220816053410/https://www.industry.gov.au/data-and-publications/australias-artificial-intelligence-action-plan>.

[81] Rishi Bommasani et al, *On the Opportunities and Risks of Foundation Models* (Centre for Research on Foundation Models, 2021).

[82] Galvagna (n 47).

[83] Ibid 5.

has outlined, there are strong reasons not to follow the European approach: the regulatory and legislative safeguards in Australia are very different and, in several key aspects, weaker. There is a pressing need to include a much wider range of social and legal expertise in any regulatory regime for AI. And precisely how to combine the requisite technical expertise to address AI's technical aspects with the social and political expertise and legitimacy required to tackle serious questions about competing rights and interests is still not clear.

In that respect, Australia faces a moment of opportunity. AI demands new regulatory approaches, new ways of facilitating deliberation and information sharing and new and better ways of deploying cross-disciplinary expertise to socio-technical problems. In building regulatory capacity for AI, the government should keep certain key priorities in view. Policymakers should take care in the delegation of regulatory discretion — and especially the allocation of decision-making power about social and political matters — to private governance bodies. Standards are likely to be better at providing assurance about the trustworthiness of processes and organisational arrangements than about the overall question of whether an AI system is safe and responsible. Standards and certification regimes can more readily provide assurance of process and technical features of AI systems than of socio-technical features; yet, even here consensus may still be out of reach and governance arrangements that are more sensitive to uncertainty and disagreement may be needed.[84] Standards are most effective if integrated well with a wider regulatory ecosystem where additional regulatory levers, including oversight and guidance by government regulators, and civil liability further direct the capabilities and incentives of participants in the AI value chain and the assurance ecosystem. [85]

It is the responsibility of government to develop, with meaningful input from civil society, guidance and frameworks for addressing the impacts of AI on individual and collective rights and interests. Efforts to bridge the expertise gap in the AI assurance expertise ecosystem, and indeed to develop multidisciplinary AI governance expertise throughout the regulatory ecosystem as a whole, deserve serious investment. Likewise, while AI standards bodies are clearly open to including a range of stakeholders and expertise, meaningful, democratic participation in standards-making and AI governance more generally will require further investment and active policy initiatives.

---

[84] Stirling (n 75).

[85] See e.g., Colin Scott (2001) Analysing regulatory space: fragmented resources and institutional design. Public Law, 329–53; Fiona Haines and Christine Parker, Reconstituting the Contemporary Corporation Through Ecologically Responsive Regulation, *Company and Securities Law Journal*, 2023, 39(6), 316–31.

Acknowledgements

This working paper is based on a submission to government made by the authors. In developing the ideas for that submission and this paper, the authors have benefited from discussions with regulators, auditors and certifiers, and other scholars at two workshops conducted in May 2023 at the University of Birmingham, UK: 'Assurance Regimes for Data-Driven Services'[86] and 'Standards and Assurance for Trustworthy Data-Driven Technologies'.[87]

[86] This workshop was hosted by Professor Karen Yeung and Dr Rotem Medzini from the University of Birmingham as part of a project led by Yeung pursuant to the European Lighthouse on Secure and Safe AI Network of Excellence www.elsa-ai.eu.The workshop, convened under Chatham House rules was attended by professionals in standards, certification and accreditation, from across Europe, with experience in assurance regimes for GDPR, medical devices and artificial intelligence.

[87] This workshop (organised by Dr Henry Fraser and Professor Christine Parker from ADM+S with the support of our colleagues at UoB) was an open international academic workshop with participants from a range of disciplines including law, computer science, political science, regulatory theory, criminology, science and technology studies, and platform governance.

admscentre.org.au

Australian Government

Australian Research Council

This Centre is funded by the Australian Government through the Australian Research Council

ADM+S