

SHOULD AUSTRALIA FOLLOW EUROPE'S APPROACH TO AI STANDARDS AND REGULATION?

HENRY FRASER, CHRISTINE PARKER, FIONA HAINES,
JOSÉ-MIGUEL BELLO Y VILLARINO, AND KIMBERLEE WEATHERALL^{*†}

This paper critically evaluates the approach in Europe's *Artificial Intelligence Act* to standards in AI regulation, and considers the suitability of 'transplanting' that approach to Australia. The AI Act uses standards to guide the implementation of legislative requirements aimed at promoting 'trustworthy AI'. As a result, standards bodies play the role of 'regulatory intermediary' (a term coined by scholars of regulatory governance such as Abbot, Snidal and Levi-Faur) interposed between government regulators and regulatory targets. We explain how Europe's use of standards for AI regulation is shaped by a set of institutional constraints and capabilities that are distinctive to the European context. Drawing on regulatory intermediary theory, we argue that the kinds of regulatory discretion that Europe's AI Act delegates to standards and assurance bodies — calling for difficult judgments about rights and the public interest — exceed their expertise and legitimacy. We identify challenges for inclusion in standard-making,

^{*} Henry Fraser (BA/LLB (Sydney), MSt, DPhil (Oxon)) is a lecturer at Queensland University of Technology Law School. This work was completed as a research fellow at the ARC Centre of Excellence for Automated Decision-Making and Society.

Christine Parker (BA, LLB, PhD) is Professor of Law at Melbourne Law School, the University of Melbourne and a Chief Investigator in the ARC Centre of Excellence for Automated Decision-Making and Society.

Fiona Haines (BA (Hons), PhD) is Professorial Fellow in the School of Social and Political Sciences at the University of Melbourne and Honorary Professor at the School of Regulation and Global Governance, ANU.

José-Miguel Bello y Villarino (Lic. Law, Lic. Pol.Sc., LL.M., LL.M., MA, PhD) is a Senior Research Fellow at the Law School of the University of Sydney and an ARC EC Industry Fellow. This work was completed as a Research Fellow at the ARC Centre of Excellence for Automated Decision Making and Society.

Kimberlee Weatherall (BA, LLB (Hons) (Sydney), BCL (Hons) (Oxon), LL.M. (Yale)) is a Professor of Law at the University of Sydney and a Chief Investigator with the ARC Centre of Excellence for Automated Decision-Making and Society

[†] This article developed from a submission to government made by the authors. In developing the ideas for that submission and this article, 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' and 'Standards and Assurance for Trustworthy Data-Driven Technologies'. The first 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. 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. The second, organised by Dr Henry Fraser and Professor Christine Parker from ADM+S (with the support of our colleagues at the University of Birmingham) 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.

and misaligned incentives that may undermine the goal of trustworthy AI (or in Australia, safe and responsible AI). Over-reliance on standards would be particularly problematic in Australia, where institutional arrangements are very different to Europe. We therefore make some suggestions as to how to make best use of standards for AI, and to avoid their pitfalls. AI standards may be useful for promoting trustworthy processes and for facilitating quantitative assessments of system inputs and outputs, including resource use. They will not, however, be well-suited for resolving difficult questions of ethics, public policy, and law, such as how to oversee and explain life-changing automated decisions. Finally, we urge regulators to prioritise efforts to develop and support the cross-disciplinary capabilities and inclusive, deliberative institutions needed to govern AI effectively.

I INTRODUCTION

The Australian Government recently accelerated its efforts to develop regulation for artificial intelligence ('AI'). The Department of Industry, Science and Resources released its Voluntary AI Safety Standard (better thought of as a government guidance document rather than a technical standard) in September 2024.¹ In the same month it conducted a consultation on 'mandatory guardrails' for AI with a proposals paper.² This followed a flurry of activity which began late in 2023. On 17 January 2024, in an interim response to its consultation on 'Safe and Responsible AI in Australia', the Government announced that it will consider 'possible legislative vehicles for introducing mandatory safety guardrails for AI in high-risk settings'.³ In the meantime, the Government committed to working with industry to develop a voluntary 'AI Safety Standard'.⁴ In February 2024, the Government announced the appointment of an Expert Group on AI, tasked with steering the development of the mandatory guardrails.⁵ Australia's new prioritisation of AI regulation followed several landmark global developments. In December 2023, the European Parliament and the Council of Europe reached provisional agreement on the final form of Europe's new regulation on artificial intelligence ('*EU AI Act*'), which takes a risk-

¹ Department of Industry, Science and Resources, *Voluntary AI Safety Standard* (Standard, 5 September 2024) ('*Voluntary Standard*').

² Department of Industry, Science and Resources, *Safe and Responsible AI in Australia: Proposals Paper for Introducing Mandatory Guardrails for AI in High-risk Settings* (Proposals Paper, September 2024) ('*Government's Proposals Paper*').

³ Department of Industry, Science and Resources, *Safe and Responsible AI in Australia Consultation: Australian Government's Interim Response* (Report, 17 January 2024) 6 ('*Government's Interim Response*').

⁴ *Ibid.*

⁵ Ministers for the Department of Industry, Science and Resources, 'New Artificial Intelligence Expert Group' (Media Release, 14 February 2024) <<https://www.minister.industry.gov.au/ministers/husic/media-releases/new-artificial-intelligence-expert-group>>.

based approach to regulation aimed at promoting ‘trustworthy’ AI.⁶ European legislators adopted the Act in June 2024.⁷ In November 2023, twenty-seven countries including Australia also committed to the development of risk-based policies on AI by signing the Bletchley Declaration on AI.⁸ Prior to that, on the 30 October 2023, the US President issued an Executive Order on the Safe, Secure and Trustworthy Development and Use of Artificial Intelligence.⁹

Whatever the final shape of any AI regulation in Australia, the general direction of development seems reasonably clear: a risk-based approach, applying guardrails in some form to uses of AI in high-risk settings. However, at the time of writing, the particular mechanisms for reaching that result remain uncertain. Regardless of the legislative mechanism, a key question will be how the (necessarily high-level) guardrails will be operationalised, including how technical standards may be incorporated. In relation to that operationalisation, the Government’s Proposals Paper mentions that the regulatory approach to technical standards in the EU AI Act could be considered in Australia.¹⁰

Australia has previously taken inspiration from European approaches to technology regulation. Australian product liability law and medical devices regulation both borrow heavily from Europe.¹¹ Planned reforms to Australia’s Privacy Act would also bring Australia more closely in line with Europe’s General Data Protection Regulation (‘GDPR’).¹² The global influence of European privacy, data and technology regulation (sometimes described as the ‘Brussels effect’) is well known.¹³ All of these factors,

⁶ ‘Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI’, *News European Parliament* (Press Release, 9 December 2023) <<https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai>>.

⁷ *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) [2024] OJ L 12/7 (‘EU AI Act’).*

⁸ See further ‘The Bletchley Declaration by Countries Attending the AI Safety Summit’, *Department of Industry, Science and Resources* (Web Page, 2 November 2023) <<https://www.industry.gov.au/publications/bletchley-declaration-countries-attending-ai-safety-summit-1-2-november-2023>>.

⁹ *Executive Order No 14110*, 88 Fed Reg 24283 (30 October 2023).

¹⁰ *Government’s Proposals Paper* (n 22) 31.

¹¹ See, eg, David Harland, ‘The Influence of European Law on Product Liability in Australia’ (1995) 17(2) *Sydney Law Review* 336; Petahn McKenna, ‘Australian Medical Device Regulations: An Overview’ in Jack Wong and Raymond Tong (eds), *Handbook of Medical Device Regulatory Affairs in Asia* 361 (Jenny Stanford Publishing, 2nd ed, 2018).

¹² Attorney-General’s Department, *Privacy Act Review Report* (Report, Australian Government, 16 February 2023) <https://www.ag.gov.au/sites/default/files/2023-02/privacy-act-review-report_0.pdf>; Australian Government, *Government Response to the Privacy Act Review Report* (Response Report, 28 September 2023) <<https://www.ag.gov.au/sites/default/files/2023-09/government-response-privacy-act-review-report.pdf>>.

¹³ See further Anu Bradford, *The Brussels Effect: How the European Union Rules the World* (Oxford University Press, 2020) (‘*The Brussels Effect*’).

taken together with the detail, scope and ambition of Europe's approach to AI regulation, may encourage Australian policymakers, including the new AI Expert Group, to take inspiration from Europe once again.

One of the defining features of the EU AI Act is the way that it uses technical standards to guide the implementation of legislative requirements for AI systems. The EU AI Act states the essential requirements for certain AI systems in a general way, and technical standards bodies then intermediate by developing more detailed standards to assist regulatory targets with compliance. Compliance with standards is not mandatory. However, conformity with approved standards creates a presumption of conformity with the Act's essential requirements.¹⁴ This approach makes standards organisations 'regulatory intermediaries' in the design and implementation of AI regulation. 'Regulatory intermediary' is a term coined by David Levi-Faur and co-authors to describe, analyse and evaluate the role of those actors that sit between official regulators and regulatory targets in the regulatory process. While regulation is often imagined as a two-party system consisting of regulators and targets, regulatory intermediary theory points out that both regulators and targets frequently lack capabilities, authority, or legitimacy needed for regulation, and therefore call on intermediaries to assist.¹⁵

Drawing on regulatory intermediary theory, this paper assesses strengths, weaknesses, challenges, and opportunities of relying on standards and assurance to manage risks from AI systems and to achieve 'responsible' and 'trustworthy' AI.¹⁶ It offers a critical evaluation of the EU AI Act's approach to regulatory intermediation in AI regulation, and considers the suitability of 'transplanting' that approach to Australia.¹⁷ Europe's use of standards for AI regulation is shaped by a set of institutional constraints and capabilities that are distinctive to the European context. We suggest that the kinds of regulatory discretion that Europe's AI Act delegates to standards and assurance bodies – calling for difficult judgments about rights and the public interest – exceed their expertise and legitimacy. The paper also identifies challenges for inclusion in standard-making, and misaligned incentives that may undermine the effectiveness of

¹⁴ *EU AI Act* (n 7) art 40.

¹⁵ See, eg, David Levi-Faur, 'Regulatory Capitalism' in Peter Drahos (ed), *Regulatory Theory: Foundations and Applications* (Australian National University Press, 2017) 289; Kenneth Abbott, David Levi-Faur and Duncan Snidal, 'Introducing Regulatory Intermediaries' (2017) 670(1) *The Annals of the American Academy of Political and Social Science* 6; Kenneth Abbott, David Levi-Faur and Duncan Snidal, 'Enriching the RIT Framework' (2017) 670(1) *The Annals of the American Academy of Political and Social Science* 280; Kenneth Abbott, David Levi-Faur and Duncan Snidal, 'Theorizing Regulatory Intermediaries: The RIT Model' (2017) 670(1) *The Annals of the American Academy of Political and Social Science* 14.

¹⁶ See, eg, Levi-Faur (n 15); Rebecca Schmidt and Colin Scott, 'Regulatory Discretion: Structuring Power in the Era of Regulatory Capitalism' (2021) 41(3) *Legal Studies* 454; Rotem Medzini and David Levi-Faur, 'Self-Governance via Intermediaries: Credibility in Three Different Modes of Governance' (2023) 25(3) *Journal of Comparative Policy Analysis: Research and Practice* 323.

¹⁷ On legal transplants, see further Toby Goldbach, 'Why Legal Transplants?' (2019) 15(1) *Annual Review of Law and Social Science* 583.

standards in promoting the EU AI Act's goal of trustworthy AI. It suggests that overreliance on standards and standards bodies as regulatory intermediaries in relation to AI would be particularly problematic in Australia, because this country lacks checks and balances on standard-setting that apply in the EU.

Nevertheless, AI standards are developing quickly and will inevitably play an important role in AI governance in Australia. Uncertain as the future of AI standards in Australia is, there is value in understanding when and how standards are likely to make the best contribution to the regulation of AI. We suggest that standards are most likely to be effective in supporting reliable processes and facilitating quantification of key inputs and outputs for AI, rather than in dealing with value-laden questions about rights and public policy. The latter are best addressed in guidance documents issued by government organisations with the legitimacy to speak authoritatively about rights and public policy. The recently published Australian Voluntary AI Safety Standard, with its 10 voluntary guardrails, is a promising example of what such guidance could look like. Whilst described as a 'standard', it is more accurately described as a set of qualitative guidelines than a technical standard. Its legitimacy could be further enhanced, however, through a more inclusive, consultative process as it is updated and developed over time: emulating, for example, the process adopted by the US Government developing its *AI Risk Management Framework* and *Risk Management Profile for Artificial Intelligence and Human Rights*.¹⁸ Whatever regulatory approach the Australian Government takes to AI, it must prioritise efforts to build the requisite governance capabilities and technical understanding across government, non-governmental organisations (including standards bodies), and civil society.

This paper proceeds in four parts. Part II explains the key features of the EU AI Act, especially the approach to standards in regulation, known as the 'New Approach', that Europe has adopted for AI. Part III sets out some key insights from theories of how regulatory intermediaries could in principle enhance trust in regulation in democracies and common dangers associated with reliance on intermediaries. Part IV applies those lessons to evaluate the suitability of the European approach for AI in general. Part V provides further reasons why Australia need not follow the European approach too closely. Recognising the significant role that standards are still likely to play in AI governance, however, Part V also offers guidance about how to make the most of them.

¹⁸ See, eg, National Institute of Standards and Technology, 'Artificial Intelligence Risk Management Framework' (2 January 2024) <<https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>> ('*NIST AI Risk Management Framework*'). For a further history of the development of the National Institute of Standards and Technology's AI Risk Management Framework, see '*AI RMF Development*', *National Institute of Standards and Technology* (Web Page, 2 January 2024) <<https://nist.gov/itl/ai-risk-management-framework/ai-rmf-development>>.

II THE EU AI ACT AND THE NEW APPROACH

In this part we explore key features of the EU AI Act. We draw attention to the way that it delegates significant regulatory and policy discretion to standards bodies in order to address a particular set of historical and institutional constraints that apply to European law and policymaking.¹⁹

The EU AI Act *aims* to promote ‘trustworthy’ AI with regulations that apply to AI horizontally across industry sectors. It takes a risk-based approach to regulation, with the burden of regulation tailored to the level of risk to rights and safety posed by different AI applications. It prohibits certain intrusive or harmful uses of AI (such as art 5(1)(c), which bans the use of general-purpose social scoring unrelated to the contexts in which the data was originally generated). A greater part of the Act deals with requirements for ‘high-risk AI systems’ that pose significant risks to health, safety or fundamental rights.²⁰ These include requirements of data governance, accuracy, human oversight, quality assurance, documentation and logging, explainability and risk management. The Act provides that conformity with harmonised standards approved by European standardisation organisations creates a *presumption of conformity* with these essential requirements for ‘high-risk’ AI systems.²¹ The Act also sets out requirements for general purpose AI models and systems (such as the well-known ChatGPT chatbot).²² The Act contemplates the development of codes of practice for providers of general-purpose AI models, under the oversight of the AI Office. The intention is that providers should be able to rely on codes of practice to demonstrate compliance with their obligations under the Act. Likewise, the AI Office can approve harmonised standards, compliance with which will also create a presumption of conformity with the Act.

The presumption of conformity based on compliance with standards follows Europe’s (nearly 40-year-old) ‘New Approach’ to technical harmonisation and standards. The New Approach, established in 1985, was designed to create a uniform approach to product safety.²³ It has applied to a wide range of products, from toys, to boats, to personal protective equipment.²⁴ It was meant to facilitate confidence and

¹⁹ We are grateful to Dr Jake Goldenfein for drawing our attention to this history, and suggesting an analysis of its influence on the *AI Act*.

²⁰ *EU AI Act* (n 7) art 6(2)–(3), annex III.

²¹ *Ibid* art 17(1)(e), 40, 42.

²² *Ibid* art 53–54.

²³ *Council Resolution of 7 May 1985 on a New Approach to Technical Harmonization and Standards* [1985] OJ C 136/1 (‘*Council Resolution on Technical Harmonization and Standards*’).

²⁴ *Ibid*; ‘New Legislative Framework’, *European Commission* (Web Page) <https://single-market-economy.ec.europa.eu/single-market/goods/new-legislative-framework_en> (‘*New Legislative Framework*’); See further Lukasz Gorywoda, ‘The New European Legislative Framework for the Marketing of Goods’ [2009] 16 *Columbia Journal of European Law* 161, 163–4.

movement of goods in the European single market. An updated ‘New Legislative Framework’, approved in 2008, cemented the New Approach, while also establishing more detailed rules about its implementation — for example regarding the accreditation of certifiers (‘notified bodies’), and requirements for certification known as ‘conformity assessment’.²⁵ Critically, this is the default technique to regulate new technological innovations and the preferred way to assess and manage their risks in Europe.²⁶

The purpose of the New Approach was to manage the challenge of ‘harmonising’ the regulation of product safety throughout Europe. The European Union (‘EU’) only has ‘competence’ (constitutional power) to make exhaustive laws that supersede the national laws of member states in certain domains. Foremost among these domains is the making of law to promote the functioning of the internal market of the EU under art 114 of the *Treaty on the Functioning of the European Union* (‘TFEU’).²⁷ Where member states’ regulation of products diverges, the flow of products in a single market will be obstructed by the cost of complying with too many different regulations. This may justify exhaustive regulation, pursuant to art 114 of the TFEU to promote the single market. This process is known as ‘harmonisation’.

The EU AI Act is squarely aimed at ‘harmonisation’ of European law on AI. Its express purpose is to create a uniform legal framework for AI, directly applicable in all member states, to improve the functioning of the internal market, while also protecting health, safety, and fundamental rights.²⁸ The Act expressly relies on art 114 of the TFEU for its validity, along with art 16 (a competence with respect to the protection of personal data).

The New Approach and New Legislative Framework place limits on the harmonisation power. Early experience with harmonisation revealed that it tended to lead to overly prescriptive regulation, and an excessive demand for uniformity across member states.²⁹ One often cited case of potential overreach

²⁵ *New Legislative Framework* (n 24).

²⁶ *Ibid.* See also ‘Evaluation of the New Legislative Framework’, *European Commission* (Web Page) <https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12654-Industrial-products-evaluation-of-the-new-legislative-framework_en>.

²⁷ *Treaty of Lisbon Amending the Treaty on European Union and the Treaty Establishing the European Community, opened for signature 13 December 2007*, [2007] OJ C 306/1 (entered into force 1 December 2009) art 114. See further Andrew McGee and Stephen Weatherill, ‘The Evolution of the Single Market: Harmonisation or Liberalisation’ (1990) 53(5) *Modern Law Review* 578.

²⁸ *EU AI Act* (n 7) recital 1.

²⁹ Piet Jan Slot, ‘Harmonisation’ (1996) 21(5) *European Law Review* 378, 381.

was the myth that diverse European food cultures would be displaced by ‘Eurosausages’ due to new EU food safety rules.³⁰ As McGhee and Weatherill put it:

‘The object of the New Approach, apart from the obvious practical point that it saves the Commission time and money, is to reconcile the diversity of cultural and commercial tradition in the Community with the need for a common Community approach in the pursuit of free trade and economic integration’.³¹

Under the New Approach, the EU avoids the ‘Eurosausage’ problem through setting rules about products at a relatively high level of abstraction, in the form of ‘essential requirements’.³² Member states and product providers are free to meet these requirements in their own way, and then mutually recognise the lawfulness of products from other states that meet the essential requirements.

Standards provide a voluntary, uniform means of demonstrating conformity with the essential requirements of regulation. That, in turn, facilitates mutual recognition of regulatory compliance among member states. 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 EU.³³ 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 requested by the European Commission, developed by European standardisation bodies and finally approved by the European Commission) is voluntary, but it creates a presumption of conformity with legislative requirements. There is therefore a strong incentive for firms to use standards to meet regulatory obligations, rather than to try to interpret and meet the more broadly drafted general requirements themselves.

Because the EU AI Act follows the New Approach and New Legislative framework logic, AI providers and other actors in AI value chains can meet their legal obligations by conducting (and in some cases perhaps self-certifying) conformity assessments against harmonised European standards.³⁴ 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.³⁵ The European Commission has already issued a

³⁰ Ibid.

³¹ McGee and Weatherill (n 2727) 584.

³² Gorywoda (n 24) 163.

³³ *Council Resolution on Technical Harmonization and Standards* (n 23) annex II.

³⁴ *EU AI Act* (n 7) art 17(1)(e), 40, 42.

³⁵ 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, 102.

standardisation request to the major European standardisation organisations, the European Committee for Standardisation ('CEN') and the European Committee for Electrotechnical Standardisation ('CENELEC').³⁶ 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 provisions, human oversight, and accuracy specifications for AI systems. Beyond this, the proposal also includes obligations to establish quality management and post-market monitoring systems and conformity assessments.³⁷ Joint Technical Committee 21 ('JTC 21'), a joint technical committee of CEN and CENELEC, is responsible for the development of these standards.³⁸ 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 under development by JTC 1-SC 42, a joint technical committee of the International Standards Organisation ('ISO') and the International Electrotechnical Commission ('IEC').³⁹ 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 between ISO and IEC.⁴⁰

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.⁴¹ European regulators, experts 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.

³⁶ *Commission Implementing Decision on a Standardisation Request to the European Committee for Standardisation and the European Committee for Electrotechnical Standardisation in Support of Union Policy on Artificial Intelligence* (European Commission Document, No C(2023)3215, 22 May 2023) <[https://ec.europa.eu/transparency/documents-register/detail?ref=C\(2023\)3215&lang=en](https://ec.europa.eu/transparency/documents-register/detail?ref=C(2023)3215&lang=en)> ('EU Commission Implementing Decision on Standardisation Request').

³⁷ Ibid 2.

³⁸ CEN and CENELEC, *Business Plan for JTC 21* (Business Plan Report, 16 March 2022) <<https://standards.cencenelec.eu/BPCEN/2916257.pdf>>.

³⁹ Ibid; JTC 1 / SC 42 has published 20 such deliverables, with more than 30 others still under development. See further, 'Standards by ISO/IEC JTC1/SC 42', *International Organization for Standardization* (Web Page) <<https://www.iso.org/committee/6794475/x/catalogue/p/0/u/1/w/0/d/0>>.

⁴⁰ Agreement on Technical Cooperation between the International Organization for Standardization and the European Committee for Standardization (20 October 2001) <[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](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)>.

⁴¹ Christine Parker, *The Open Corporation: Effective Self-Regulation and Democracy* (Cambridge University Press, 2002).

Manufacturers (or in the case of the EU AI Act, AI providers), who best understand the conditions on the ground, take primary responsibility for conformity assessment.⁴²

In practice, however, the line between technical implementation and policymaking is always blurred.⁴³ The development of the network protocols underlying the internet, for example, baked in a set of political and social values about the terms on which information is accessed and exchanged.⁴⁴ The TCP/IP protocol promotes ‘openness’ in the exchange of information over security or privacy: a value choice that ultimately facilitates now-commonplace extractive data practices by large technology companies.⁴⁵ That is not to say that openness is not an important value – only that apparently technical choices may settle trade-offs between competing values in consequential ways.

This blurring of boundaries between the technical and the socio-political is particularly pronounced in the EU AI Act. ‘High-risk AI systems’ and general-purpose AI systems are not like toys or boats or even medical devices. High-risk systems, as defined by the Act, include systems used to make decisions about people and their interests, in law enforcement, university admissions, job recruitment and essential services.⁴⁶ They may also include systems the failure of which could cause significant environmental harm, such as where AI is used to manage flows of water or electricity through access networks or to monitor weather conditions to avoid floods or fires. These AI systems are socio-technical

⁴² *Council Resolution on Technical Harmonization and Standards* (n 23) annex 2; See also, Commission of the European Communities, *Communication from the Commission to the Council and the European Parliament: Enhancing the Implementation of the New Approach Directives* (European Commission Document, No COM(2003) 240, 7 May 2003) pt 1.3 <<https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2003:0240:FIN:en:PDF>>.

⁴³ See, eg, Sandra Braman, ‘The Framing Years: Policy Fundamentals in the Internet Design Process, 1969–1979’ (2011) 27(5) *The Information Society* 295, 296; Corinne Cath and Luciano Floridi, ‘The Design of the Internet’s Architecture by the Internet Engineering Task Force and Human Rights’ (2017) 23(2) *Science and Engineering Ethics* 449, 453; Michael Veale, Kira Matus and Robert Gorwa, ‘AI and Global Governance: Modalities, Rationales, Tensions’ [2023] 19 *Annual Review of Law and Social Science* 255, 261; Merijn Chamon, *The European Parliament and Delegated Legislation: An Institutional Balance Perspective* (Bloomsbury Publishing, 2022); José-Miguel Bello y Villarino, ‘Global Standard-Setting for Artificial Intelligence: Para-Regulating International Law for AI?’ (2023) 41(1) *Australian Year Book of International Law* 157, 159; Hans-W Micklitz, *The Role of Standards in Future EU Digital Policy Legislation: A Consumer Perspective* (Report, July 2023) 1 <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 16).

⁴⁴ Lawrence Lessig, *Code and Other Laws of Cyberspace* (Basic Books, 1999); Braman (n 43) 20.

⁴⁵ Andrew L Russell, *Open Standards and the Digital Age: History, Ideology, and Networks* (Cambridge University Press, 2014); See also, Lorie Merson, ‘The Net Has Never Been Neutral’, *loriemerson* (Blog Post, 14 August 2021) <<https://loriemerson.net/2021/08/14/the-net-has-never-been-neutral/>>. Merson points out that less lofty concerns, such as the interests of a developing computer industry using interoperability of communication between computers to avoid standardisation of computer software and hardware, may have been the most influential factors shaping the development of internet architecture.

⁴⁶ *EU AI Act* (n 7) annex III: See the relevant high-risk AI Systems referred to in art 6(1)–(2).

systems. Their risk–cost–benefit profile is complex and polyvalent. Whether such systems are responsible or trustworthy depends not only on how they are developed and deployed, but also on how they are used, by whom, to what end, and in what context.⁴⁷

Because the Act sets out its essential requirements at a high level 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.⁴⁸ There are bounds to this discretion, but control over standardisation by the European governing bodies (or member states) tends to be indirect and attenuated. The terms of the European Commission’s standardisation request on AI are, if anything, even more general than the essential requirements in the Act.⁴⁹ The European Commission’s representatives participate in technical committees as observers, where they may 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.⁵⁰ Having drafted the essential requirements of the EU 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.⁵¹ In doing so, it can signal its regulatory preferences. However, by their nature, objections are piecemeal rather than being vehicles for coherent, carefully expounded policy.

The EU AI Act also delegates other forms of regulatory discretion to other regulatory agencies. Member state government agencies designated as ‘market surveillance authorities’ have powers to monitor and audit AI systems already on the market.⁵² There are also procedures for these authorities to evaluate,⁵³ and impose additional requirements on, systems which are compliant with the Act’s requirements

⁴⁷ Roel Dobbe, ‘System Safety and Artificial Intelligence’ in Justin Bullock et al (eds), *The Oxford Handbook of AI Governance* (Oxford University Press, 1st ed, 2024).

⁴⁸ Veale and Borgesius (n 35).

⁴⁹ *EU Commission Implementing Decision on Standardisation Request* (n 36).

⁵⁰ *Regulation (EU) No 1025/2012 of the European Parliament and of the Council of 25 October 2012 on European standardisation, amending Council Directives 89/686/EEC and 93/15/EEC and Directives 94/9/EC, 94/25/EC, 95/16/EC, 97/23/EC, 98/34/EC, 2004/22/EC, 2007/23/EC, 2009/23/EC and 2009/105/EC of the European Parliament and of the Council and repealing Council Decision 87/95/EEC and Decision No 1673/2006/EC of the European Parliament and of the Council* [2012] OJ L 316/12, art 10 (5)–(6) (*‘Regulation on European Standardisation’*).

⁵¹ *Ibid* art 11.

⁵² *EU AI Act* (n 7) art 74.

⁵³ *Ibid* art 79.

(including through conformity with standards).⁵⁴ There are mechanisms for members of the public to complain to these authorities, with agencies directed to take such complaints into account in conducting their market surveillance.⁵⁵ The European Commission has the power to issue guidelines on the practical implementation of the Act's requirements, including requirements for high-risk AI systems.⁵⁶ The Act also establishes an Artificial Intelligence Board and AI Office, which have wide advisory and coordination functions.⁵⁷ Most of these provisions were introduced by amendments proposed by the European Parliament and the Council of the European Union to the original draft of the Act.⁵⁸ The amendments were made with the benefit of highly critical feedback on the Commission's original proposal for the Act and the role of standards.⁵⁹ They ensure that the application and development of standards occurs under the shadow of potential intervention by government agencies, and with the possible assistance of guidance from the AI Office, the AI Board and European Commission. This arrangement may enable government bodies to exercise indirect influence over the standards process.

Nevertheless, 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, a large share of control over those matters is delegated to standards bodies. The European Commission cannot use observers, adoption or objections to supplant the standards process.⁶⁰ The European Parliament also has no binding veto over harmonised standards mandated by the Commission.⁶¹ And even interventions by market authorities or the AI Office do not necessarily mean standards will be changed. The influence of these other regulatory actors on standards, or on the approach to certification against standards, is indirect. In short, the price of harmonising European AI law is the

⁵⁴ Ibid art 82.

⁵⁵ Ibid art 85.

⁵⁶ Ibid art 96.

⁵⁷ Ibid ch VII.

⁵⁸ *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 - General Approach* (Proposal No 2021/0106(COD), 25 November 2022) <<https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/en/pdf>>; *Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts* (Adopted Amendments No P9_TA(2023)0236, 14 June 2023) <https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html> ('*Amendments Adopted by EU Parliament on Harmonised Rules on AI*').

⁵⁹ See, eg, Veale and Borgesius (n 35).

⁶⁰ Redeker Sellner Dahs, *The European System of Harmonised Standards: Q&A Commissioned by the German Federal Ministry for Economic Affairs and Energy* (Response Report, August 2020).

⁶¹ Veale and Borgesius (n 35); *Regulation on European Standardisation* (n 50) art 11.

relinquishment of a considerable amount of control by government over highly charged, socio-political details of its implementation.

III STANDARDS FOR AI: CHALLENGES OF EXPERTISE, LEGITIMACY, INCLUSION AND INCENTIVES

Trustworthy AI is the stated goal of the EU AI Act. Safe and responsible AI is the theme of the consultation process regarding AI regulation in Australia. In this part, we assess the suitability of standards to promote safe, responsible, trustworthy AI. We consider this question, at first, on its own terms: is an arrangement such as the one in Europe well adapted to promote trustworthiness? We reserve for the next part consideration of how the differences between the Australian and European context may impact the suitability and desirability of transplanting the European approach to AI regulation.

Drawing on the theory of regulatory intermediaries, we suggest that it is, at best, unclear whether standards bodies and standards are well suited to the task of regulating AI systems, especially high-risk AI systems with the potential for significant impacts on important legal and human rights. We briefly introduce regulatory intermediary theory, deriving four criteria that are helpful for evaluating the European approach to AI regulation and standards, and highlighting the challenges of using standards to regulate AI. We then assess the European approach against those criteria.

A *Regulatory Intermediaries*

In the European regime just described, standards bodies can be understood as ‘regulatory intermediaries’. They are third parties that sit between official regulators and regulatory targets and that draw on their own capabilities, authority and legitimacy to assist both regulators and targets.⁶² David Levi-Faur and co-authors have suggested that the enrolment of regulatory intermediaries, such as bodies that set and certify to technical standards, auditors and other professionals, into the regulatory process can help build trust in regulation by avoiding both over- and under- regulation.⁶³ The theory of regulatory intermediaries emerges from a tradition in regulation and governance that recognises that effective and legitimate regulation – systematically and reliably influencing behaviour of targets – cannot be achieved

⁶² Abbott, Levi-Faur and Snidal, ‘Introducing Regulatory Intermediaries’ (n 15); Medzini and Levi-Faur (n 16).

⁶³ Abbott, Levi-Faur and Snidal, ‘Introducing Regulatory Intermediaries’ (n 15) 7; Luc Bres, Sébastien Mena and Marie-Laure Salles-Djelic, ‘Exploring the Formal and Informal Roles of Regulatory Intermediaries in Transnational Multistakeholder Regulation’ (2019) 13(2) *Regulation & Governance* 125, 128; Medzini and Levi-Faur (n 16) 325.

merely through heavily-enforced top-down rules, but nor is pure self-regulation likely to be effective or legitimate. Rather, regulation requires communication and feedback between regulators and targets, and is typically enhanced through the involvement of third parties in assisting this process.⁶⁴

At their best, regulatory intermediaries help to implement rules, translating them into practical forms useful to regulatory targets, and absorbing, processing and translating feedback from targets back to regulators.⁶⁵ They are most effective when they have the capacity to interpret existing rules, audit, certify and supervise with greater competence and legitimacy (and at lower cost) than government regulators.⁶⁶ However they, like any other participant in regulation, may lose credibility through regulatory capture.⁶⁷ And the rule of law may be undermined if too much regulatory discretion and power over important matters of public policy is arrogated to unaccountable private actors.⁶⁸ The enrolment of regulatory intermediaries (including standards bodies in the regulation of AI) should therefore be attuned to these strengths and weaknesses both in general terms and as they relate to a specific region or jurisdiction.

We draw from the literature four key (overlapping) criteria for evaluating the trustworthiness, as regulatory intermediaries, of standardisation and assurance bodies involved in AI regulation.⁶⁹ While the criteria we identify are not the only determinants of trustworthiness, they are widely accepted in normative commentary on regulatory intermediaries, and indeed commentary on regulation in general.

Firstly, regulatory intermediaries must have sufficient capacity and expertise in the area in which they exercise regulatory discretion.⁷⁰ Ideally, they should have superior competence to conventional regulators, at least in their particular domain. At best they will have sufficient expertise to transfer

⁶⁴ See, eg, Ian Ayres and John Braithwaite, *Responsive Regulation: Transcending the Deregulation Debate* (Oxford University Press, 1992) 54; Julia Black, 'Critical Reflections on Regulation' [2002] 27 *Australian Journal of Legal Philosophy* 1; Christine Parker, 'Twenty Years of Responsive Regulation: An Appreciation and Appraisal' (2013) 7(1) *Regulation & Governance* 2.

⁶⁵ Abbott, Levi-Faur and Snidal, 'Introducing Regulatory Intermediaries' (n 15), 8; Bres, Mena and Salles-Djelic (n 63) 128; Graeme Auld and Stefan Renckens, 'Rule-Making Feedbacks through Intermediation and Evaluation in Transnational Private Governance' (2017) 670(1) *The Annals of the American Academy of Political and Social Science* 93.

⁶⁶ Medzini and Levi-Faur (n 16) 339.

⁶⁷ Ibid; Abbott, Levi-Faur and Snidal, 'Enriching the RIT Framework' (n 15), 285.

⁶⁸ Jan Freigang, 'Scrutiny: Is Responsive Regulation Compatible with the Rule of Law?' (2002) 8(4) *European Public Law* 463; Leighton McDonald, 'The Rule of Law in the "New Regulatory State"' (2004) 33(3) *Common Law World Review* 197; Schmidt and Scott (n 16).

⁶⁹ *EU AI Act* (n 7) recital 1. See also, Department of Industry, Science and Resources, *Safe and Responsible AI in Australia: Discussion Paper* (Discussion Paper, Australian Government, June 2023) 3 <https://storage.googleapis.com/converlens-au-industry/industry/p/prj2452c8e24d7a400c72429/public_assets/Safe-and-responsible-AI-in-Australia-discussion-paper.pdf>.

⁷⁰ Abbott, Levi-Faur and Snidal, 'Enriching the RIT Framework' (n 15); Abbott, Levi-Faur and Snidal, 'Introducing Regulatory Intermediaries' (n 15).

knowledge to regulatory targets through a process of education, persuasion, and capacity-building.⁷¹ Even if they are enrolled for reasons other than expertise (such as cost-saving or efficiency), they must not be, or be perceived to be, incompetent.⁷²

Secondly, regulatory intermediaries will not promote trust in themselves or in AI unless their exercise of power is, and is perceived to be, legitimate.⁷³ Indeed, one of the main reasons to enrol them may be to *enhance* the legitimacy of a regulatory framework and the feasibility of implementing details by ensuring relevant industry and stakeholder experience and expertise is utilised in implementing regulation.⁷⁴

Thirdly, the inclusiveness or degree of democracy in decision-making is related to legitimacy, but important enough in its own right to count as an independent criterion for evaluation.⁷⁵ The structures of membership and decision-making naturally impact on the trustworthiness of regulatory intermediary arrangements, and building communities of trust is a key reason to enrol a regulatory intermediary in the first place.⁷⁶ It is for this reason that multi-stakeholder consultation is enshrined in the ISO/IEC guide on standardisation.⁷⁷

Fourthly, the incentives of regulatory intermediaries should be aligned with the objectives of regulation and the relevant regulatory agency.⁷⁸ This requires a degree of independence from targets, some resilience against regulatory capture, and the avoidance of clear conflicts of interest. As laws and institutions differ between jurisdictions, the extent to which apparently similar regulatory intermediaries, or intermediary arrangements, meet these criteria may vary substantially.

B *Expertise and legitimacy*

⁷¹ Jeroen Van Der Heijden, 'Brighter and Darker Sides of Intermediation: Target-Oriented and Self-Interested Intermediaries in the Regulatory Governance of Buildings' (2017) 670(1) *The Annals of the American Academy of Political and Social Science* 207, 210.

⁷² Abbott, Levi-Faur and Snidal, 'Enriching the RIT Framework' (n 15).

⁷³ Schmidt and Scott (n 16); Abbott, Levi-Faur and Snidal, 'Introducing Regulatory Intermediaries' (n 15); Bres, Mena and Salles-Djelic (n 63); Karen Lee, 'Legitimacy in the New Regulatory State' (PhD Thesis, University of New South Wales, March 2016).

⁷⁴ Abbott, Levi-Faur and Snidal, 'Introducing Regulatory Intermediaries' (n 15).

⁷⁵ Ibid; Schmidt and Scott (n 16) 459.

⁷⁶ Abbott, Levi-Faur and Snidal 'Introducing Regulatory Intermediaries' (n 15) 7.

⁷⁷ International Standards Organization and International Electrotechnical Commission, *Recommended Practices for Standardization by National Bodies* (Formal Standards No 59, 2019).

⁷⁸ See generally Abbott, Levi-Faur and Snidal, 'Introducing Regulatory Intermediaries' (n 15); Abbott, Levi-Faur and Snidal, 'Enriching the RIT Framework' (n 15).

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. Evaluating the acceptability of risks to fundamental rights engages difficult questions of politics, public policy, law, and ethics, and has generally been the purview first of policymakers, and of courts.⁷⁹ It is not clear that standards bodies have the right combination of expertise and political legitimacy to credibly determine the value-laden questions of public policy that arise in the governance of high-risk AI systems that impact fundamental rights.⁸⁰ It is telling, for instance, that the European Commission's Joint Research Centre concluded 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 begin its own work on an ambitious 'Checklist for AI Risk Management'.⁸¹

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. Expertise with determining when a product is safe, for instance, does not necessarily equip standards bodies to design processes and methods that determine (or at least heavily influence) how accuracy in predictions of criminal recidivism must be determined; 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 life-changing automated decisions to different stakeholders such as affected individuals, courts and regulators.

Compare the challenge of ensuring a toy is safe and trustworthy to the challenge of ensuring the safety and trustworthiness of a high-risk AI system such as the controversial Suspect Targeting Management Plan, used by the New South Wales Police Force to identify targets for ongoing police

⁷⁹ Ranj Zuhdi and Hadrien Pouget, 'AI and Product Safety Standards Under the EU AI Act', *Carnegie Endowment for International Peace* (Web Page, 5 March 2024) <<https://carnegieendowment.org/2024/03/05/ai-and-product-safety-standards-under-eu-ai-act-pub-91870>>.

⁸⁰ Veale and Borgesius (n 35); Karen Yeung and Nathalie Smuha, 'Operationalising Trustworthy AI Governance: Beyond Motherhood and Apple Pie?' (Working Paper, ADM+S Symposium, 20 July 2022).

⁸¹ Josep Soler Garrido et al, *Analysis of the Preliminary AI Standardisation Work Plan in Support of the AI Act* (JRC Technical Report No 132833, 17 May 2023) <<https://publications.jrc.ec.europa.eu/repository/handle/JRC132833>>.

monitoring and intervention.⁸² For (traditional) toys, safety concerns are of course very important, but the domain of risk is relatively narrow, confined mostly to physical risks such as choking, strangulation, or risks of children hurting others with the toy. For the Suspect Targeting Management Plan, the considerations which may be simultaneously in play include not only safety and cost, but also privacy; non-discrimination; personal liberty; freedom of association; compliance with laws regarding police power; and the effective administration of law enforcement and public order. Judgments about the appropriate degree and kind of accuracy, explainability, data governance, and risk management for a high-risk system will often engage competing rights of different individuals and groups, as well as public interests, such as interests in innovation, or in efficiencies generated by the use of AI. And yet it is for these thorny judgments that technical standards are supposed to be used – in Europe at least. Even if standards bodies did have the depth and breadth of expertise to engage confidently with these kinds of questions, it is extremely difficult to address risks as complex and intangible as discrimination and invasion of privacy in precise, standardised ways.⁸³

Issues of political legitimacy also arise.⁸⁴ 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?⁸⁵ In Europe, this degree of discretion seems inconsistent with the division of responsibilities envisioned in the New Approach. It seems to impinge into the domain of formulating essential regulatory requirements rather than purely technical matters of implementation.

C Inclusion

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.⁸⁶

⁸² Vicky Sentas and Camilla Pandolfini, *Policing Young People in NSW: A Study of the Suspect Targeting Management Plan* (Report, 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/>>; See also, Law Enforcement Conduct Commission, *An Investigation into the use of the NSW Police Force Suspect Targeting Management Plan on Children and Young People* (Final Report, October 2023) <<https://www.lecc.nsw.gov.au/news-and-publications/publications/operation-tepito-final-report.pdf/@download/file>>.

⁸³ Zuhdi and Pouget (n 79).

⁸⁴ See generally Rotem Medzini and Karen Yeung, 'Background Paper: Assurance Regimes for Data-Informed Services' (22 May 2022); Yeung and Smuha (n 80); 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) 15 *European Journal of Risk Regulation* 431.

⁸⁵ Schmidt and Scott (n 16); Veale and Borgesius (n 35).

⁸⁶ See generally Julia Black, *Rules and Regulators* (Clarendon Press, 1997).

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 (what is a good explanation of an AI decision? How should a principle of fairness be implemented?).⁸⁷ 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.⁸⁸ 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 and ideally should be involved throughout the whole lifecycle of those instruments, from their inception to their regular review.⁸⁹

Generally neither creation of, nor access to, standards is sufficiently inclusive.⁹⁰ The Regulation on European Standardisation, the Commission's draft standardisation request on AI and the European Parliament's proposed amendments to the EU AI Act all emphasise the need to promote stakeholder participation in standards-making, including participation by civil society organisations.⁹¹ Standards Australia has also done much 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, including academics, lawyers, policy experts, and members of civil society organisations. In the EU, for example, a specific

⁸⁷ For further discussion on the ambiguity and uncertainty associated with these questions, see 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; Maria Nordström, 'AI under Great Uncertainty: Implications and Decision Strategies for Public Policy' (2022) 37(4) *AI & Society* 1703; Lindsay Weinberg, 'Rethinking Fairness: An Interdisciplinary Survey of Critiques of Hegemonic ML Fairness Approaches' [2022] 74 *Journal of Artificial Intelligence Research* 75.

⁸⁸ Schmidt and Scott (n 16) 467.

⁸⁹ Ibid; Lee (n 73).

⁹⁰ Christine Galvagna, *Inclusive AI Governance* (Discussion Paper, 30 March 2023) <<https://www.adalovelaceinstitute.org/report/inclusive-ai-governance/>>; Gillian Hadfield and Jack Clark, 'Regulatory Markets: The Future of AI Governance' (ArXiv No 2304.04914, 25 April 2023).

⁹¹ *Regulation on European Standardisation* (n 50); *Amendments Adopted by EU Parliament on Harmonised Rules on AI* (n 58) amendments 103 and 104.

participatory mechanism is ongoing at the time of writing for the Code of Practice for general purpose AI described above.⁹²

However, without dedicated and sustained funding and support for this kind of participation, it is unclear how long this mix of expertise can be maintained.⁹³ The language of the Australian Government's Interim Response to the safe and responsible AI consultation also suggests a less inclusive development process, where the CSIRO's NAIC 'work[ed] with industry' to develop the ten voluntary AI safety guardrails discussed in the introduction to this paper.⁹⁴ Based on the published acknowledgments, the development of the Voluntary Safety Standard was informed by input from a range of invited entities including government agencies and regulators. It is not clear, however, how these entities were chosen to provide input. They appear, for the most part, to have already belonged to the NAIC's 'Responsible AI Network', rather than having been invited in a more open process. Only one civil society group (Choice) was involved, along with the Diversity Council of Australia.⁹⁵ To maximise the legitimacy (and usefulness) of that document, and of any mandatory guardrails document that it develops in future, the Government should commit to an inclusive process of consultation which includes civil society, academia and affected stakeholders for any further updates or iterations of the guidance.

Barriers to participation in the development of technical standards also tend to be practical rather than formal. Joining a committee to work on a given standard (eg, 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 civil society organisations and small- and medium-sized businesses do not.⁹⁶ As a consequence, large commercial interests tend to have a disproportionate influence over the development and content of standards.

There is also a problem of access to standards which fits broadly under the umbrella of concerns about inclusion, but which has implications for the rule of law. It costs nothing to access legislation,

⁹² 'First Draft of the General-Purpose AI Code of Practice published, written by independent experts', *European Commission* (Press Release, 14 November 2024) <<https://digital-strategy.ec.europa.eu/en/library/first-draft-general-purpose-ai-code-practice-published-written-independent-experts>>.

⁹³ Camille Dornier, *For a 'Standardisation Governance Act' - ANEC and BEUC Recommendations to Adapt Regulation (EU) 1025/2012* (Recommendations Report, 23 January 2024) <https://www.beuc.eu/sites/default/files/publications/BEUC-X-2024001_For_a_standardisation_governance_act.pdf>.

⁹⁴ *Government's Interim Response* (n 3) 6.

⁹⁵ See 'Acknowledgements' section in *Voluntary Standard* (n 1) 58.

⁹⁶ Galvagna (n 90) 42. See also Micklitz (n 43); Henk 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* (Report, Erasmus University, August 2009) <https://www.erim.eur.nl/fileadmin/default/content/erim/content_area/news/2009/smeaccessreport%202009.pdf>.

although it can often be challenging for ordinary citizens to find and interpret the relevant law. But access to standards comes with a price tag. Standards Australia charges approximately AUD 814 per year for package subscriptions to standards in particular fields such as construction.⁹⁷ Individual ISO/IEC standards deliverables such as technical standards and technical reports cost upwards of CHF 100 (AUD 173), with the technical report on bias in AI systems, for example, currently priced at CHF 166 (AUD 294).⁹⁸ 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, including most civil society organisations.⁹⁹ Some standards, such as the Institute of Electrical and Electronics Engineers' ('IEEE') AI standards are free, but it is not clear whether these will create a presumption of conformity in the European context, nor whether there is any role for them in Australia's current wave of AI policymaking.

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.¹⁰⁰ 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 able to know without first paying substantial sums of money, is inconsistent with this basic tenet.

D *Risk of misaligned incentives*

Regulatory intermediaries always have their own substantive and organisational objectives.¹⁰¹ As with all forms of regulation and governance, standards and assurance processes are susceptible to capture and misaligned incentives.¹⁰² Standards bodies and big audit firms have an interest in privatising regulation, and in making it complex, in order to expand the market for their services and the services of standards professionals.¹⁰³ Industry participants in standards-making have an interest in including their own practices

⁹⁷ 'Curated Subscriptions', *Standards Australia* (Web Page) <<https://store.standards.org.au/sets>>.

⁹⁸ See 'Standards by ISO/IEC JTC 1/SC 42', *International Standardization Organisation* (Web Page) <<https://www.iso.org/committee/6794475/x/catalogue/p/1/u/1/w/0/d/0>>.

⁹⁹ On cost as a barrier to access to standards, see, eg, de Vries et al (n 96).

¹⁰⁰ 'Rule of Law', *Parliamentary Education Office* (Web Page) <<https://peo.gov.au/understand-our-parliament/how-parliament-works/system-of-government/rule-of-law/>>.

¹⁰¹ Abbott, Levi-Faur and Snidal, 'Introducing Regulatory Intermediaries' (n 15).

¹⁰² Abbott, Levi-Faur and Snidal, 'Enriching the RIT Framework' (n 15) 285. See generally, Daniel Carpenter and David Moss, *Preventing Regulatory Capture: Special Interest Influence and How to Limit It* (Cambridge University Press, 2013).

¹⁰³ Abbott, Levi-Faur and Snidal, 'Enriching the RIT Framework' (n 15), 285; Luc Fransen and Genevieve LeBaron, 'Big Audit Firms as Regulatory Intermediaries in Transnational Labor Governance' (2019) 13(2) *Regulation & Governance* 260, 262.

and preferences into standards, lowering their own costs and raising the costs of their competitors.¹⁰⁴ It is natural for companies that have invested in a certain way of doing business to be resistant to costly changes, and to want to keep the barriers to market entry up. Here we come to a point of overlap with the analysis on inclusion. Incumbent companies such as we have just described might respond to our concerns about access to standards and the rule of law along the following lines: if smaller AI developers do not have sufficient investment to be able to afford access to standards and professional advice as to how to implement them, then they should not be considered as trustworthy to enter the market. But to take that position would be to permit expensive technical standards to operate as a form of lock in. Standards could entrench the position of industry incumbents, and exclude new market entrants who might have better, more innovative and more diverse approaches to various kinds of AI applications.

Potentially misaligned incentives do not just apply to powerful industry incumbents. Any business using standards for self-certification (which the EU AI Act contemplates) may have conflicts between their commercial interests and the public interests supposed to be protected through the certification process. The enrolment of third-party certifiers as yet another group of regulatory intermediaries may address this problem somewhat, but 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.¹⁰⁵ 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.¹⁰⁶ 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.¹⁰⁷ All of this suggests that, to the extent AI standards play a role in Australian AI regulation, duties of certifiers need to be more clearly specified. It may also be necessary to create additional duties to report to government about risks that fall outside the scope of certification.

¹⁰⁴ Maurits Dolmans, 'Standards for Standards European Union Law' (2002) 26(1) *Fordham International Law Journal* 163, 171.

¹⁰⁵ 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.

¹⁰⁶ See, eg, *Ku-ring-gai Council v Chan* [2017] NSWCA 226 ('*Ku-ring-gai*').

¹⁰⁷ See, eg, 'More for Show than Safety: Certificates in the Textile Industry', *European Centre for Constitutional and Human Rights* (Web Page) <<https://www.ecchr.eu/en/case/more-for-show-than-safety-certificates-in-the-textile-industry/>>.

IV HOW SHOULD AUSTRALIA APPROACH STANDARDS FOR AI?

In this part, we consider the implications of the foregoing analysis for the development and adoption of AI standards in Australia. We marshal several reasons not to adopt the European approach to AI regulation as a ‘regulatory transplant’ in Australia. Some of the problems of the regulatory intermediary arrangements under the EU AI Act might be worsened if that model were adopted in Australia due to the different legal and institutional structures in place.

Nonetheless, it is clear that standards will play a role in the governance of AI both internationally and in Australia. The Government released its Voluntary Safety Standard (as we’ve noted, a document more like a qualitative guide) in September.¹⁰⁸ ISO/IEC has published 20 standards on AI already, with more than 30 currently under development.¹⁰⁹ IEEE has a broad portfolio, which includes publicly accessible foundational AI standards.¹¹⁰ 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.¹¹¹ We therefore make suggestions here as to when standards, and regulatory intermediation by standards bodies, may be most useful and effective in AI regulation. We also offer some reflections on the questions must be answered, and what measures implemented to make the most of them.

A Risks of a ‘regulatory transplant’

The analysis above suggests that there are several reasons not to adopt the European approach to AI standards wholesale as a regulatory transplant. Using standards to fill in the most important, value-laden, politically charged details of AI regulation, and deeming conformity with such standards to achieve compliance with regulation may be ill-advised because of the problems of expertise, legitimacy, inclusion and incentives described above. It is yet to be seen whether standards are up to the task set for them by Europe’s AI Act. AI governance, and especially governance of systems that sort and prioritise people’s access to important services and entitlements, is inherently multidisciplinary and involves difficult socio-technical questions. The legitimacy of standards bodies may come into question if standards are expected to resolve civic questions about rights, public policy, and competing public interests. There are serious

¹⁰⁸ *Voluntary Standard* (n 1); See also *Government’s Interim Response* (n 3).

¹⁰⁹ International Standardization Organisation ‘Standards by ISO/IEC JTC 1/SC 42’ (Web Page) <<https://www.iso.org/committee/6794475/x/catalogue/p/0/u/1/w/0/d/0>>.

¹¹⁰ Institute of Electrical and Electronics Engineers Standards Association, ‘Autonomous and Intelligent System Standards’ (Web Page) <<https://standards.ieee.org/initiatives/autonomous-intelligence-systems/standards/>>.

¹¹¹ Bonnie Osif, ‘Make It Safe and Legal’ in Michael Fosmire and David Radcliffe (eds), *Integrating Information into the Engineering Design Process* (Purdue University Press, 2014) 115.

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.

Even if ‘AI safety’ standards remain voluntary here in Australia, their capacity to exert ‘soft’ regulatory pressure on a wide range of policy issues will mean that expertise and legitimacy remain pressing issues. The nomenclature that the Australian Government has adopted in requesting the development of standards – of an ‘AI safety’ standard – is therefore somewhat concerning. ‘Safety’ evokes the concept of physical safety or product safety, with an attendant sense of binarity between the safe and unsafe, and an emphasis on technical reliability.¹¹² It obscures the socio-technical dimensions of AI risks discussed at length above. Reframing the project in more open ended terms – responsible AI guidelines, for example – would better reflect the challenges involved.¹¹³ The Australian Government’s choice to call upon the National AI Centre (‘NAIC’), a group located at the time within Australia’s national science agency, the CSIRO, to develop the standard is also interesting.¹¹⁴ It apparently circumvents the already-running Standards Australia process, with all the efforts at inclusion, multi-disciplinarity, and expertise-building that have been involved.

On the positive side, the Voluntary AI Safety Standard much more closely resembles a set of government guidelines than a true technical standard. It is much more oriented around qualitative social and political matters than qualitative technical ones, and it was developed by a government entity (rather than a private regulatory intermediary such as ISO). Its guidance is nuanced and helpful, and was informed by consultation not only with industry, government, and other entities, but also with Australia’s ‘AI expert group’ which included lawyers and social science academics as well as specialists in machine learning.

However, its legitimacy would have been further enhanced if the process of development had been more inclusive still, bringing in the perspectives and expertise of a wider range of stakeholders beyond academic experts. In particular, more opportunities for a wider range of civil society and consumer organisations, and for persons at risk of being affected by AI systems, to contribute meaningfully would have been valuable. Rural and regional Australians, Indigenous Australians and other marginalised groups tend to be at greatest risk of harm from the use of AI, and ought therefore to be

¹¹² Roel Dobbe, “‘Safety Washing’ at the AI Safety Summit” (LinkedIn, 10 November 2023) <<https://www.linkedin.com/pulse/safety-washing-ai-summit-roel-dobbe-gv4oe/>>.

¹¹³ We are grateful to our colleagues at the Gradient Institute for this observation.

¹¹⁴ NAIC is now part of Australia’s Department of Industry, Science and Resources, but was outside that department at the time most of this work happened.

involved in its regulation. We strongly urge the Government to pursue this increase in inclusion both in further updates to the voluntary standard, and in developing its proposed ‘mandatory guardrails’.

Questions of legitimacy and inclusion in standards-making are especially urgent in Australia, because Australia does not have the same institutional arrangements around the enrolment of standards bodies as exist in Europe. 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.¹¹⁵

Europe also has explicit, formal protection of human rights; for instance, in the *Charter of Fundamental Rights of the European Union*.¹¹⁶ 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 identifies a system as ‘high-risk’ and thus required to conform with essential requirements. Europe’s consumer protection law is also arguably more developed in areas of importance to the regulation of private sector use of AI: Europe has, for example, established prohibitions on unfair commercial practices that Australia lacks;¹¹⁷ updated privacy legislation (the GDPR) has been in place since 2018 and new regulations have been developed for digital services.¹¹⁸ Australia has 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 and enforcing human rights and consumer rights 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 cooperation throughout the regulatory, standardisation and assurance ecosystems; regulators and standard-setting are connected.¹¹⁹

¹¹⁵ Goldbach (n 17).

¹¹⁶ *Charter of Fundamental Rights of the European Union* [2000] OJ C 364/1.

¹¹⁷ *Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market and amending Council Directive 84/450/EEC, Directives 97/7/EC, 98/27/EC and 2002/65/EC of the European Parliament and of the Council and Regulation (EC) No 2006/2004 of the European Parliament and of the Council (Unfair Commercial Practices Directive)* [2005] OJ L 149/22, 22–39.

¹¹⁸ *Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act)* [2022] OJ L 277/1.

¹¹⁹ For a snapshot of key institutions and instruments, see *European Commission*, ‘Vademecum on European Standardisation’, (Web Page) <https://single-market-economy.ec.europa.eu/single-market/european-standards/vademecum-european-standardisation_en>.

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.¹²⁰ 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 might provide false reassurance, conferring unwarranted consumer confidence but without preventing significant failure.

There are, however, advantages to the differences in institutional arrangements between Australia and Europe. The Australian Government is not subject to the same combination of institutional and constitutional constraints that pushed the European Commission to adopt the New Approach in regulating AI. Australia is in a position to take the best from European regulation, and to dispense with the parts that do not suit its regulatory goals.

And importantly, some of the basic outlines of the EU AI Act *are* worth emulating. Other jurisdictions such as the US and Canada have taken up the European blueprint of a 'risk-based' approach, where heavier regulatory burden falls upon systems that pose greater risk. Likewise, there appears to be general alignment between the EU, US and Canada on the basic *kinds* of requirements for high-risk systems and generative AI models, especially systems used in government.¹²¹ These include requirements for impact assessment, explainability, oversight, documentation, data governance, testing, validation and monitoring, and risk assessment and mitigation. Aligning the fundamentals of Australian AI regulation with the baseline set by these international regulatory frameworks would be sensible and would help Australian AI providers participate in the global market.

But Australia need not follow Europe in giving so much regulatory discretion over policy questions to standards bodies. It should recognise that there are particular domains where standardisation should not be used to underpin, or stand in for legislation, government policy, or other forms of civic consensus-building: especially in the definition of ethical values, important rights, or difficult questions of public interest. Leading European consumer organisations have gone so far as to recommend European legislation enshrining such a principle.¹²²

¹²⁰ Regarding cooperation in Australia's assurance infrastructure, see, eg, Department of Industry Innovation and Science, *Best Practice Guide to Using Standards and Risk Assessments in Policy and Regulation* (Report, July 2016) 13 <<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>>.

¹²¹ Alignment between the EU, US, UK, and several other nations was recently formalised in the European Framework Convention on Artificial Intelligence.

¹²² Dornier (n 93).

B *When standards work best*

Because of the differences between the European and Australian regulatory context, Australian policymakers can and should use standards and standards bodies (and other forms of co-regulation) in ways that better play to the strengths of regulatory intermediation, and standards bodies as regulatory intermediaries. Standards should primarily cover areas in which the expertise of the regulatory intermediary — standards bodies — is strongest. Enrolling standards bodies in domains where their expertise and credibility are strongest enhances, rather than detracts from, the legitimacy of intermediation.

Standards overwhelmingly deal with specifications, procedures, and guidelines aimed at promoting safety, consistency, and reliability.¹²³ It is in relation to these qualities and arrangements — safety, consistency, reliability, and process — that standards bodies' expertise is most valuable, and standards are likely to work best.¹²⁴ Standards bodies will best operate as regulatory intermediaries 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.¹²⁵ 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

¹²³ See Jose-Miguel Bello y Villarino et al, *Standardisation, Trust and Democratic Principles: The Global Race to Regulate Artificial Intelligence* (Report, United States Studies Centre, 31 July 2023) <<https://www.ussc.edu.au/standardisation-trust-and-democratic-principles-the-global-race-to-regulate-artificial-intelligence>>.

¹²⁴ Fraser and Bello y Villarino (n 84).

¹²⁵ Department of Industry, Innovation and Science, *Best Practice Guide to Using Standards and Risk Assessments in Policy and Regulation* (Report, July 2016) 6 <<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>>.

place, with responsibilities for AI risks allocated to appropriately senior and qualified people within an organisation.¹²⁶

One emerging area of AI governance where standards may play an important role is in the domain of environmental protection and sustainability. Environmental impacts have been something of an afterthought in ‘responsible AI’ discourse and governance.¹²⁷ Yet, the extensive environmental impact of AI systems and their supply chains is striking. As a result, responsible AI experts are increasingly calling for governance measures to encourage sustainable AI.¹²⁸ 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 for materials to be used in the production of necessary equipment like graphics processing units, and the waste created by the regular upgrading of equipment to keep up with expanding AI development and applications.¹²⁹

Serious consideration of the environmental impact of AI was a relatively late inclusion in the negotiation of the EU AI Act. After representations by NGOs and Green Party members,¹³⁰ the European Parliament’s June 2023 draft of the Act recognised ‘environmental friendliness’ as a priority of AI regulation and importantly included provisions that would require providers of AI systems deemed to be high risk to produce an environmental risk assessment and to make use of appropriate standards to reduce environmental impact, particularly energy use and carbon emissions.¹³¹ The final version of the Act somewhat watered down these provisions. The purposes of the Act still include ensuring ‘environmental protection’ as an element of the overall goal of ‘ensuring a high level of protection of health, safety [and]

¹²⁶ *NIST AI Risk Management Framework* (n 18) 8.

¹²⁷ Roberto Verdecchia, June Sallou, Luiz Cruz, ‘A Systematic Review of Green AI’ (2023) 13(4) *WIREs Data Mining and Knowledge Discovery* 1.

¹²⁸ See, eg, Philipp Hacker, ‘Sustainable AI Regulation’ (2024) 61(2) *Common Market Law Review* 345; Bogdana Rakova and Roel Dobbe, ‘Algorithms as Social-Ecological-Technological Systems: An Environmental Justice Lens on Algorithmic Audits’ (Conference Paper, ACM Conference on Fairness, Accountability, and Transparency, 2023) 491.

¹²⁹ See, eg, Sasha Luccioni, ‘The mounting human and environmental costs of generative AI’, *Ars Technica* (Web Page, 4 December 2023) <<https://arstechnica.com/gadgets/2023/04/generative-ai-is-cool-but-lets-not-forget-its-human-and-environmental-costs/>>; Anne-Laure Ligozat et al, ‘Unraveling the Hidden Environmental Impacts of AI Solutions for Environment Life Cycle Assessment of AI Solutions’ (2022) 14(9) *Sustainability* 5172.

¹³⁰ AlgorithmWatch, ‘EU Artificial Intelligence Act – Recommendations on Ecological Transparency’ (Report, April 2023) <<https://algorithmwatch.org/en/wp-content/uploads/2023/05/2023-05-02-EU-Artificial-Intelligence-Act---recommendations-on-ecological-transparency.pdf>>.

¹³¹ *Draft Compromise Amendments on the Draft Report Proposal for a regulation of the European Parliament and of the Council on harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts* (Draft Amendments, Committee on the Internal Market and Consumer Protection Committee on Civil Liberties, Justice and Home Affairs, 5 June 2023) art 1, 9(2)(a), 28b(2)(d) <https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/CJ40/DV/2023/05-11/ConsolidatedCA_IMCOLIBE_AI_ACT_EN.pdf>. For a summary see Hacker (n 128) 371–4.

fundamental rights’ in the uptake of AI.¹³² The focus however is on voluntary codes of conduct for assessing and minimising environmental impact.¹³³ A set of draft provisions that explicitly required the creation of standards to quantify, log, and make transparent energy consumption and other environmental impacts of AI systems were not included in the final version of the Act. Instead, the Act makes more generically worded statements that high-risk AI systems must conduct risk assessments and create technical documentation processes that address their impact on ‘health, safety [and] fundamental rights’, which (as noted above) is defined to include environmental protection.¹³⁴

There is an opportunity for standards to operate as a helpful form of regulatory intermediation in promoting sustainable AI. If Australia chooses to require environmental impact assessments for AI systems, or even general impact assessments that include sustainability, standards have the potential to assist regulatory targets in quantifying energy and water use, and other environmental impacts of AI systems.¹³⁵ Water, energy, mineral and land use lend themselves more easily to quantification than do human rights risks and impacts - although difficult choices must always be made about which dimensions of the various environmental impacts of any new technology can and should be measured, and in which material context and making what assumptions about the sourcing of energy, material and water.¹³⁶ Technical standards about practices, processes and design requirements all have the potential to provide genuinely useful guidance to targets about how to detect and reduce environmental impact. Additionally, there is rich social science literature and community practice setting out both the limitations of quantification and the potential for more contextual, inclusive and diverse reporting and accounting for environmental impacts and imagining new ways to pursue sustainability.¹³⁷ There is already a well-established ESG ecosystem where standards play a role in guiding environmental impact assessments, and certifying sustainability practices, across a range of industries.¹³⁸ Thus in this context, the problem of expertise, and therefore also to some extent legitimacy, is less pronounced.

¹³² *EU AI Act* (n 7) art 1.

¹³³ *Ibid* art 95(2), 112(7).

¹³⁴ *Ibid* art 9 and 11 read with art 1 and annex IV. See also, Hacker (n 128) 371–4.

¹³⁵ See further OECD, ‘Measuring the Environmental Impacts of Artificial Intelligence Compute and Applications: The AI Footprint’ (Working Paper No 341, November 2022) <<https://www.oecd.org/publications/measuring-the-environmental-impacts-of-artificial-intelligence-compute-and-applications-7babf571-en.htm>>.

¹³⁶ Anne Pasek, Hunter Vaughan and Nicole Starosielski, ‘The World Wide Web of Carbon: Toward a Relational Footprinting of Information and Communications Technology’s Climate Impacts’ (2023) 10(1) *Big Data and Society* 1.

¹³⁷ *Ibid*; Rakova and Dobbe (n 128); Melissa Gregg and Yolande Strengers, ‘Getting Beyond Net Zero Dashboards in the Information Technology Sector’ [2024] (February) 108 *Energy Research and Social Science* 2214.

¹³⁸ Hacker (n 128).

Naturally, there will still be challenges. Choices which appear to be technical always have normative implications. For example, in conversations about AI impact, the environmental impact of any end use of AI is still almost entirely out of scope, as is the underlying business model of more AI embedded in more products and services meaning greater energy and resource consumption.¹³⁹ For instance, an environmentally-friendly AI may be used to identify new fossil fuel mining sites or to program and personalise advertising to promote unnecessary consumption.¹⁴⁰ This wider-reaching kind of analysis did not seem to be in contemplation in earlier drafts of the EU AI Act that contained more detailed environmental protection provisions. With any practice- or process-based standard, questions may still remain about whether process and practices reliably produce the more substantive goal (such as trustworthiness or fairness) that is intended; and indeed, substantive goals may be contested.¹⁴¹ This is why standards-making and regulatory intermediation must be supported with appropriate institutional arrangements.

C *Institutional arrangements: question and priorities*

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? Further research and policy discussions are needed to develop a clearer sense of the role of standards bodies, industry bodies and other stakeholders, 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.

As noted above, the response to negative feedback about earlier drafts of the EU AI Act was to make greater provision for government agencies (such as the AI Office, AI Board, and market surveillance authorities) to develop guidance on the implementation of the Act's more value-laden requirements; and

¹³⁹ Gregg and Strengers (n 137).

¹⁴⁰ 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' (2023) 36 *Philosophy & Technology* 25:1–34 1. See also T Donaghy, C Henderson and E Jardim, *Oil in the Cloud: How Tech Companies are Helping Big Oil Profit From Climate Destruction* (Report, Greenpeace, 19 May 2020) <<https://www.greenpeace.org/usa/oil-in-the-cloud/>>.

¹⁴¹ Kira Matus and Michael Veale, 'Certification Systems for Machine Learning: Lessons from Sustainability' (2022) 16(1) *Regulation & Governance* 177.

to have the final word on whether some systems pose risks which are unacceptable, even when they conform with standards.

Government should have a key role not only in establishing red lines regarding unacceptable uses of AI, but also in developing guidance to assist targets to meet regulatory objectives such as ‘trustworthiness’ or ‘responsible’ practice. 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. Government may need to provide additional guidance to stakeholders in the standards-making process on difficult public policy questions. And, where an international standard or part of a standard does not match Australian requirements, the Government may specifically issue guidance to this effect. The European Commission does so with technical standards.¹⁴² European consumer organisations have suggested the development of additional institutional arrangements, such as assessment of standards by Europe’s regulatory scrutiny board, to ensure that delegations of regulatory discretion to standards bodies remain within proper bounds.¹⁴³

Australia does not have these kinds of institutions. Still, whatever government agency ends up with responsibility for ‘responsible AI’ (whether a dedicated regulator, an AI Safety Commissioner without enforcement powers, or domain-specific regulators),¹⁴⁴ guidance on difficult socio-technical questions involved in AI governance and commentary on technical standards and other forms of self-regulation should be part of its (or their) remit(s). And if the combination of expertise and legitimacy to issue this kind of guidance and commentary does not yet exist in Australia’s regulatory ecosystem, this would be a strong reason to establish a dedicated AI regulator.¹⁴⁵ Some trade-offs involved in AI development and deployment 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

¹⁴² For an account of one particular objection to a technical standard issued by the European Commission see Fraser and Bello y Villarino (n 84) pt III.

¹⁴³ Dornier (n 93).

¹⁴⁴ Australian Human Rights Commission, *Human Rights and Technology* (Final Report, 2021) 196 <<https://humanrights.gov.au/our-work/technology-and-human-rights/projects/final-report-human-rights-and-technology>>.

¹⁴⁵ See *ibid* on the recommendation to establish an AI Safety Commissioner.

to truly reckon with whether a system is trustworthy in all the circumstances.¹⁴⁶ Again, it is government that is best placed to take responsibility for facilitating this kind of deliberative, open-ended public policy.¹⁴⁷

Inclusion in the process of rulemaking is critical to ensure that deliberation captures the tensions and complexities just described, and also to ensure that the process of decision-making around controversial issues bears the stamp of legitimacy. At the level of standards-making, inclusion 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, more diverse, participatory standards-making bodies that draw from different pools of experience and expertise, and break existing moulds.¹⁴⁸ 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.¹⁴⁹

Legitimacy and inclusion in AI governance could also be enhanced by investing actively in capacity-building, and perhaps even by enrolling other kinds of regulatory intermediaries or co-regulators to assist in enriching governance discourse. The Government has hitherto emphasised the building of technical and economic capacity to take advantage of opportunities presented by AI.¹⁵⁰ 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 not only for accreditation bodies, certifiers and others involved in AI governance, but also for industry and for civil society. 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

¹⁴⁶ 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).

¹⁴⁷ Parker (n 64). See also Henry 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' (SSRN Scholarly Paper No 3960461, 30 September 2021) pt 5 <<http://dx.doi.org/10.2139/ssrn.3960461>>.

¹⁴⁸ Galvagna (n 90).

¹⁴⁹ Ibid 5.

¹⁵⁰ See, eg, Department of Industry, Science and Technology, *Australia's Artificial Intelligence Action Plan* (Report, June 2021) <<https://webarchive.nla.gov.au/awa/20220816053410/https://www.industry.gov.au/data-and-publications/australias-artificial-intelligence-action-plan>>.

likely to be valuable in capacity-building and in inclusive deliberation about governance.¹⁵¹ Australia's AI Expert Group is a promising example of the recruitment of experts from academia and civil society to assist in the development of AI regulation. This may be helpful at the outset in developing a regulatory agenda for AI in Australia in a short time frame. But ultimately, the legitimacy and effectiveness of AI regulation will depend on bringing to bear the widest range of expertise, experiences and perspectives possible, and on including representatives from the groups most likely to be affected by the use of artificial intelligence and automation.

The development of Australia's mandatory 'guardrails' for high-risk uses of AI should build on the strengths of the voluntary guardrails and aim to fill those gaps in the voluntary safety standard identified above. While there may be a role for technical standards in implementing certain guardrails focused on technical risk-management processes, the guardrails should also include, or be supplemented by, social technical guidance. This guidance should come from a body with an imprimatur of legitimacy. The process of developing that guidance should recruit participants from a wide range of backgrounds and expertise, including academia, civil society, industry and from the citizenry who are most likely to be affected by high-risk AI applications.

The Government may also wish to look further afield for exemplars. Key guidance materials from the United States Government, such as the National Institute for Standards and Technology's ('NIST') *AI Risk Management Framework*, and the US Department of State's *Risk Management Profile for Artificial Intelligence and Human Rights* provide promising examples of standard-like documents that meet challenge of both providing technical direction, and nuanced socio-technical guidance.¹⁵² For instance, NIST's AI Risk Management Framework, like Australia's voluntary guardrails, is less like a technical standard than a government guidance document. It was developed through extensive consultation with a wide range of stakeholders, and its guidance is supplemented with citations to relevant literature. In that way, it acknowledges the potential open-endedness of some AI governance questions, while also providing direction and assistance to stakeholders as to how they might approach such questions.

The institutions that created these instruments also have an appropriate level of legitimacy and authority to deal with social and political questions. The role of NIST as a regulatory intermediary is

¹⁵¹ Rishi Bommasani et al, *On the Opportunities and Risks of Foundation Models* (Research Report, Centre for Research on Foundation Models, 2021) <<https://crfm.stanford.edu/assets/report.pdf>>.

¹⁵² *NIST AI Risk Management Framework* (n 18); US Bureau of Cyberspace and Digital Policy, *Risk Management Profile for Artificial Intelligence and Human Rights* (Release, 25 July 2024) <<https://www.state.gov/risk-management-profile-for-ai-and-human-rights/>>.

particularly interesting. It is part of the United States Department of Commerce, with a role which encompasses the development of technical materials and guidance, as well as systemic research and evaluation of technologies.¹⁵³ As an institution, it combines the technical expertise of a private standards body, with the legitimacy and authority of government. But more typical command and control measures are taken by the Department of Commerce, rather than by NIST, maintaining a separation of regulatory functions.

As the Australian Government further develops our regulatory frameworks and guidance for AI, it may wish to carefully consider whether some similar institutional arrangement may be helpful in Australia. It may not wish to copy wholesale the arrangements in Europe or the US, but it should take note of where legitimacy and competency align well in those jurisdiction's approaches to AI regulation and regulatory intermediation. Finally, accountability is also likely to be important in avoiding some of the pitfalls of standards and certification, to the extent they play a role in Australian AI governance. 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 certified products or services fail.¹⁵⁴ 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 to purchasers of a property to take reasonable care to prevent loss to the purchasers as a result of 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'.¹⁵⁵ 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. Should certification play an important role in AI governance, measures for promoting accountability among participants in the standards and assurance ecosystem may be necessary.

¹⁵³ National Institute of Standards and Technology, 'About NIST' (Web Page, 11 January 2022) <<https://www.nist.gov/about-nist>>.

¹⁵⁴ Jan de Bruyne, 'Third Party Certification and Artificial Intelligence', 67–88 in Ignas Kalpkas and Julija Kalpokienė (eds), *Intelligent and Autonomous: Transforming Values in the Face of Technology* (Brill, 2023).

¹⁵⁵ *Ku-ring-gai* (n 106) [83] (Meagher JA).

V 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 EU AI Act envisions, Australia should be cautious about adopting the European approach to standards for AI. The European approach has been criticised for example in its lack of consensus on key issues, its potential human rights impact and its missed opportunity to enact more substantial measures to address environmental impact. In addition, the regulatory and legislative safeguards in Australia are very different and, in several key aspects, weaker. Both in general and specifically for Australia there is a pressing and ongoing need to include a wide 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. Yet even here consensus may still be out of reach and governance arrangements that are more sensitive to uncertainty and disagreement may be needed.¹⁵⁶ Standards are only effective if integrated 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.¹⁵⁷

It is the responsibility of government to develop guidance and frameworks to address the impacts of AI on individual and collective rights and interests, with meaningful input from civil society (especially representatives of those most likely to be affected by the use of AI). While recent initiatives such as the Voluntary AI Standard show promise, there is room for improvement in matters of inclusion and legitimacy. Efforts to bridge the expertise gap in the AI assurance expertise ecosystem, and indeed to

¹⁵⁶ See generally Stirling (n 146).

¹⁵⁷ See, eg, Colin Scott, 'Analysing Regulatory Space: Fragmented Resources and Institutional Design' [2001] (Summer) *Public Law* 283; Fiona Haines and Christine Parker, 'Reconstituting the Contemporary Corporation Through Ecologically Responsive Regulation' (2023) 39(6) *Company and Securities Law Journal* 316.

develop multidisciplinary AI governance expertise throughout the regulatory ecosystem as a whole, deserve serious investment. Likewise, while AI standards bodies and organisations such as Australia's NAIC 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.