ARE WE REGULATING THE RIGHT DIGITAL SYSTEMS? TESTING EMERGING ARTIFICIAL INTELLIGENCE FRAMEWORKS AGAINST REAL-WORLD PUBLIC SECTOR SYSTEMS

JOSÉ-MIGUEL BELLO Y VILLARINO,* KIMBERLEE WEATHERALL,** TERRY CARNEY,*** ALEXANDRA SINCLAIR**** AND SCARLET WILCOCK*****

This article critically examines the sufficiency of legal frameworks governing artificial intelligence ('AI') and automated decision-making ('ADM') systems in the public sector through doctrinal and empirical analysis, using data from 163 ADM systems in New South Wales ('NSW') government agencies. Considering the regulatory frameworks created by the European Union's ('EU') AI Act and the NSW AI Assessment Framework ('AIAF'), the article reveals that restricting regulatory oversight to AI systems neglects many high-risk technologies. Notably, three-fifths of ADM systems deemed high-risk under the AIAF do not satisfy the EU's AI definition, thereby potentially bypassing mandatory assessment. This finding challenges the assumption that AI-focused regulation adequately captures technological risk in the public sector, exposing significant regulatory blind spots. Comparative analysis further highlights interpretive and operational ambiguities within existing frameworks. The authors advocate for a risk-based regulatory approach encompassing all ADM systems, providing critical insights for Australian policymakers and enriching the discourse on responsible digital governance.

* Senior Research Fellow, The University of Sydney Law School.
** Professor, The University of Sydney Law School and Chief Investigator, ADM+S.
*** Professor Emeritus, The University of Sydney Law School and Associate Investigator, ADM+S.
**** Postdoctoral Research Fellow, The University of Sydney Law School and ADM+S.
***** Senior Lecturer, UNSW Law & Justice and Associate Investigator, ADM+S.
This research has been supported by funding from the Australian Research Council ('ARC'), and conducted by researchers from the ARC Centre of Excellence on Automated Decision-Making and Society ('ADM+S') (CE200100005). It also takes advantage of a dataset constructed as part of a university project funded and supported by the NSW Ombudsman under a research agreement entered between the NSW Ombudsman's Office and the University of Sydney on behalf of a consortium of university partner institutions of ADM+S. The authors acknowledge the work of Paul Henman, Rita Matulionyte, Lyndal Sleep, Melanie Trezise, Jenny Van Der Arend, and Sergio Pinzón in that project.

I INTRODUCTION

Public sector digital systems can harm people and societies if not developed, deployed, and monitored with care. The risks arising from automation and the use of artificial intelligence ('AI') in the public sector are well-known and rehearsed in the literature.[1] In response, governments around the world have enacted, or are in the process of considering, a range of frameworks, regulations, laws, and standards intended to ensure that systems used in the public sector which pose risks of harm to individuals or important societal interests are subject to additional governance in the process of design and development, and ongoing deployment and use.[2] Most of these emerging regulatory frameworks focus on the risks of harm posed by AI. But how well do these new regulatory frameworks capture systems of concern?

In this article, we explore a critical question: to which systems should we apply legal frameworks designed to regulate public sector automated and AI systems that impact the public? In particular, our concern is to test the appropriateness of limiting the scope of application of these frameworks to systems involving the use of AI, the definition of which is explored below. This inquiry is pursued first via a doctrinal analysis of the scope of two important examples of regulatory frameworks that apply to public sector automated decision-making ('ADM') systems and AI: the European Union's ('EU') *EU AI Act*[3] and 'The NSW AI Assessment Framework' ('NSW AIAF').[4] We then test these two frameworks against real-world systems in development or in use in government. Specifically, we draw on a unique dataset of more than 200 ADM systems in development or use by New South Wales ('NSW') government departments and agencies in early 2024, as developed in a 2023–24 project mapping ADM in the NSW public sector commissioned by the NSW Ombudsman's Office ('NSWOO').[5] The public version of that dataset with some modifications was published by the NSW Ombudsman in March 2024, alongside a detailed research report.[6]

---

1 See, eg, Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (St Martin's Press, 2018); Janina Boughey and Katie Miller (eds), *The Automated State: Implications, Opportunities and Challenges for Public Law* (Federation Press, 2021).

2 In this article we use the terminology of 'regulatory frameworks': an umbrella term intended to refer to a wide range of instruments used by policymakers to impose governance on technical systems. These range from binding legislation and regulations through to binding policies and guiding frameworks.

3 *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 2024/1689 ('*EU AI Act*').

4 NSW Government, 'The NSW Artificial Intelligence Assessment Framework' (Framework, 2024) <https://www.digital.nsw.gov.au/policy/artificial-intelligence/nsw-artificial-intelligence-assessment-framework> ('NSW AIAF'). At the time of writing the NSW AIAF is published as both an Excel spreadsheet and PowerPoint presentation. Citations in this article are to the PowerPoint presentation.

5 Kimberlee Weatherall et al, *Automated Decision-Making in NSW: Mapping and Analysis of the Use of ADM Systems by State and Local Governments* (Research Report, 8 March 2024) <https://apo.org.au/node/325901>.

6 NSW Ombudsman, *Compendium of ADM Systems* (Report, 8 March 2024) <https://cmsassets.ombo.nsw.gov.au/assets/Reports/Compendium-of-ADM-Systems.pdf> ('*Compendium of ADM Systems*').

The findings of our analysis indicate that regulators will not adequately mitigate the risks associated with technology adoption in the public sector by only regulating systems involving the use of AI. Despite its popularity with policymakers, we show that AI is a poor proxy for risks from technology adoption. This is an unsurprising conclusion: it is now well-recognised that simple automated systems can cause harm, a point patently illustrated by the Robodebt scandal,[7] the NSW Treasury's garnishee system investigated by the NSWOO,[8] and the Horizon Post Office scandal in the United Kingdom ('UK').[9] Our analysis enriches the picture by assessing and comparing these two frameworks against a large set of real-world public sector digital systems. This analysis highlights the number and proportion of those systems which are likely to fall within the areas or operations of high or elevated risk, but which would not be targeted by regulatory frameworks focused only on AI. Ultimately, our findings add significant weight to the conclusion that legal frameworks seeking to prevent and mitigate the harms associated with public sector digital systems should not be limited to AI.

A further contribution of this article lies in the analysis of the scope of the regulatory frameworks and the difficulties in their application. Specifically, we critically query the respective definitions relating to what counts as 'AI', and the classification of 'high-risk' systems in the *EU AI Act* and the NSW AIAF. This analysis reveals that the definitions adopted in these frameworks are not straightforward to apply. Examining these definitions against real world examples of digital systems throws textual uncertainties in both frameworks into sharp relief.

Whilst this research is of general relevance to legal and policy debates about the appropriate regulation of digital systems in government, it is particularly important and timely for the Australian context. At the time of writing, regulation in Australia at the federal level of public (and private) sector development and use of AI and ADM remains unsettled. The Commonwealth government has not announced a final position on the introduction of mandatory guardrails or any other AI regulation, whether generally or within government specifically; the question of governing ADM is also open following initial consultations led by the Attorney-General's Department[10] and an ongoing pilot for an Australian Government-wide AI

---

7 *Royal Commission into the Robodebt Scheme* (Report, 7 July 2023) vol 1 ('*Royal Commission into Robodebt*').

8 NSW Ombudsman, *Revenue NSW: The Lawfulness of Its Garnishee Order Process* (Report, 30 April 2024) ('*Revenue NSW Garnishee Order Report*').

9 Nick Wallis, *The Great Post Office Scandal: The Fight to Expose a Multimillion Pound IT Disaster Which Put Innocent People in Jail* (Bath Publishing, 2021). In other well-known cases, such as the Childcare Benefits Scandal in the Netherlands, the algorithms were advanced, including learning capabilities: see, eg, Lucas Haitsma and Maarten Bouwmeester, 'Learning from Control Deficits in the Childcare Benefits Scandal: A Plea for Multi-Level Analysis in Law and Policy Research' (2023) 44(3) *Recht der Werkelijkheid* 57 <https://doi.org/10.5553/RdW/138064242023044003004>.

10 Attorney-General's Department (Cth), 'Use of Automated Decision-Making by Government' (Consultation Paper, November 2024) <https://consultations.ag.gov.au/integrity/adm/user_uploads/consultation-paper-use-of-automated-decision-making-by-government.pdf> ('Use of ADM by Government').

assurance framework.[11] As such, the proper application of regulatory frameworks seeking to regulate digital systems in the public sector is an open question. We consider the implications of our research findings for the Australian context in the conclusion of this article.

This article first charts the direction of AI regulation in Part II, before moving to an examination of the *EU AI Act* and the NSW AIAF in Part III. This includes a close analysis of the definitional criteria of both AI and high or elevated risk in each framework. Part IV introduces the empirical component of the research involving the assessment of the use cases in our empirical dataset against the scope of the two frameworks. Finally, we consider the implications of this research for the application of legal frameworks seeking to mitigate the risks associated with technology adoption in government.

II REGULATION OF AI AND AUTOMATED SYSTEMS IN THE PUBLIC SECTOR CONTEXT

Over the last five years or so, governments around the world have introduced, or are in the process of developing, a range of regulatory frameworks and approaches in an effort to meet the emerging challenges associated with the adoption of AI and automation, especially in the public sector. Whilst this regulatory landscape is heterogeneous and rapidly evolving, a number of distinct approaches can be discerned. The first involves the development of administrative norms that specifically address the use of AI or ADM systems in those public law contexts. Examples of this approach can be seen in the Canadian Government's *Directive on Automated Decision-Making*,[12] 'The NSW AI Assessment [previously 'Assurance'] Framework',[13] and the now-defunct *Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* United States ('US') executive order.[14] A second approach targets the deployment of AI in pre-defined use cases, but develops

11 'DTA Pilots New AI Assurance Framework', *Digital Transformation Agency* (Blog Post, 21 October 2024) <https://www.dta.gov.au/blogs/dta-pilots-new-ai-assurance-framework>.

12 Jacqueline McIlroy, Sara Luck and Henry Fraser, 'Decoding Canada's Directive on Automated Decision-Making: A Blueprint for AI "Guardrails"?', *Medium* (Blog Post, 24 May 2024) <https://medium.com/automated-decision-making-and-society/decoding-canadas-directive-on-automated-decision-making-08124bcdf250>; 'Directive on Automated Decision-Making', *Government of Canada* (Web Page, 24 June 2025) <https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592> ('Directive on Automated Decision-Making').

13 'NSW AIAF' (n 4).

14 *Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*, Executive Order No 14110, 88 Fed Reg 75,191 (30 October 2023); repealed by *Removing Barriers to American Leadership in Artificial Intelligence*, Executive Order No 14179, 90 Fed Reg 8,741 (23 January 2025). However, the principles of its execution mechanism have remained substantially in place in the new approach: Office of Management and Budget (US), 'Driving Efficient Acquisition of Artificial Intelligence in Government' (Memorandum No M-25-22, 3 April 2025) 1–2 <https://www.whitehouse.gov/wp-content/uploads/2025/02/M-25-22-Driving-Efficient-Acquisition-of-Artificial-Intelligence-in-Government.pdf>; Office of Management and Budget (US), 'Advancing the Responsible Acquisition of Artificial Intelligence in Government' (Memorandum No M-24-18, 24 September 2024) 1–2 <https://www.whitehouse.gov/wp-content/uploads/2024/10/M-24-18-AI-Acquisition-Memorandum.pdf>.

obligations of risk assessment and mitigation irrespective of the private or public nature of the developer or deployer. The best-known example is the *EU AI Act*,[15] but this approach is also reflected in South Korea's recently enacted legislation ('*Korean Law*').[16] This 'horizontal' model of general application can be contrasted with what might be called a third approach: the 'vertical' model. This last approach seeks to amend or adapt existing domain- or sector-specific regulatory frameworks, and has been advocated, for example, by the previous government in the UK, but never materialised.[17]

Whilst the nature, stringency, and effectiveness of obligations attached to these frameworks are critical questions, and have already attracted significant academic commentary,[18] the concern of this article is the question of *scope*: that is, which types of technological systems fall within the scope of these frameworks, and which systems *ought* to. The existing frameworks are cast in various ways. Some regimes, like the *EU AI Act* or the *Korean Law*, apply to both public and private sector AI, but only in some pre-identified contexts,[19] whilst other models apply to all public sector uses, such as the NSW AIAF.[20] In other jurisdictions, such as Canada, AI is only regulated as part of an ADM system intended to be used by its federal government. Notwithstanding this variation, the majority of these regulatory frameworks are animated by two specific criteria: (1) they apply to systems involving the use of a system identified as 'AI', and (2) they are animated by categories of risk so that higher-risk systems are subject to more intensive regulatory control. These criteria are examined in turn below.

To date, almost all of the existing or proposed regulatory systems apply *only* where AI is involved. That is, 'involves AI' is effectively used as a proxy for 'gives rise to elevated risks to people'. Canada's 'Directive on Automated Decision-Making' is the standout exception: it applies to all automated decision-making across the Canadian federal government.[21] Whether confining regulatory regimes to AI makes sense – that is, whether, tested against real-world systems, a focus on AI captures all the higher risk systems warranting scrutiny – is the question at the heart of this article. For now, it is sufficient to note that to fall within the regulatory frameworks tested here – the *EU AI Act* and the NSW AIAF – a system

15 *EU AI Act* (n 3) art 2.

16 인공지능 발전과 신뢰 기반 조성 등에 관한 기본법 [The Framework Act on the Development of Artificial Intelligence and Establishment of Trust Foundation], (South Korea) Law No 20676, 21 January 2025 <https://archive.is/Mrtst> ('*Korean Law*').

17 Department for Science, Innovation & Technology (UK), 'A Pro-innovation Approach to AI Regulation' (White Paper, 29 March 2023) [47] <https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper>.

18 See, eg, the discussion in section 6 of Miguel Ángel Presno Linera and Anne Meuwese, 'Regulating AI from Europe: A Joint Analysis of the AI Act and the Framework Convention on AI' (2025) *The Theory and Practice of Legislation* (advance) <https://doi.org/10.1080/20508840.2025.2492524>.

19 See above nn 15–16.

20 NSW AIAF (n 4) 6.

21 'Directive on Automated Decision-Making' (n 12) cl 5.1, stating that '[t]his directive applies to any automated decision system used to make an administrative decision or a related assessment about a client', although note that under clause 5.2, systems being tested are not included.

must involve the use of AI. How that concept is understood in these frameworks is examined in Part III(A).

The second major concept which defines the scope of these frameworks is *risk*; that is, the level of risk of harm posed by the systems to the public. Many of these frameworks, including the two examined in this article, reflect a risk-based approach in which regulatory controls are graduated according to the level of risk posed by the system, with the highest-risk systems subject to the strictest requirements.[22] The underlying goal of risk-based approaches of this kind is striking an appropriate balance between enabling innovation and protecting the public from the potential harms of new technologies.[23] This reflects the assumption that some AI systems deserve more attention owing to their expected risk of harm. In focusing on systems of high or elevated risk, these regulatory frameworks reflect policymakers' assumption that some systems pose a lower risk of harm and hence do not require advanced scrutiny.[24]

To our knowledge, the template for this approach was first developed in Canada in its directive on ADM systems in government,[25] but it gained broader traction for AI since the 2021 proposal for an AI regulation in the EU.[26] Today it is reflected globally in the *Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*,[27] and locally in Australia's *National Framework for the Assurance of Artificial Intelligence in Government*,[28] as well as Commonwealth government proposals published in September 2024 for mandatory guardrails for high-risk AI systems.[29] In each case, more stringent or extensive obligations are imposed on the developers and deployers of higher-risk systems, although the specific obligations vary in each framework.

---

22 Robert Baldwin, Martin Cave and Martin Lodge, *Understanding Regulation: Theory, Strategy, and Practice* (Oxford University Press, 2nd ed, 2011) ch 6 <https://doi.org/10.1093/acprof:osobl/9780199576081.001.0001>.

23 Henry Fraser and José-Miguel Bello y Villarino, 'Acceptable Risks in Europe's Proposed AI Act: Reasonableness and Other Principles for Deciding How Much Risk Management Is Enough' (2024) 15(2) *European Journal of Risk Regulation* 431, 434 <https://doi.org/10.1017/err.2023.57>.

24 See, eg, Annex III of the *EU AI Act* (n 3) or systems for deployment in already heavily regulated contexts that are also considered high-risk in the *EU AI Act* as per the 'harmonisation legislation' in Annex I.

25 McIlroy, Luck and Fraser (n 12).

26 *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* [2021] COM(2021) 206 ('*EU AI Act (Proposal)*').

27 *Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*, opened for signature 5 September 2024, CETS No 225 (not yet in force) ch VE.

28 Australian Government et al, *National Framework for the Assurance of Artificial Intelligence in Government*' (Framework, 21 June 2024) 8 <https://www.finance.gov.au/sites/default/files/2024-06/National-framework-for-the-assurance-of-AI-in-government.pdf >.

29 Department of Industry, Science and Resources (Cth), *Safe and Responsible AI in Australia: Proposals Paper for Introducing Mandatory Guardrails for AI in High-Risk Settings* (Proposals Paper, September 2024) 15–17 <https://storage.googleapis.com/converlens-au-industry/industry/p/prj2f6f02ebfe6a8190c7bdc/page/proposals_paper_for_introducing_mandatory_guardrails_for_ai_in_high_risk_settings.pdf> ('Safe and Responsible AI in Australia').

III COMPARING THE *EU AI ACT* AND THE NSW AIAF

The *EU AI Act* and the NSW AIAF represent useful points of comparison to explore the question of the scope of application of frameworks seeking to regulate advanced technological systems in government. The *EU AI Act* represents a global benchmark in the AI regulatory space, whilst the NSW AIAF is the relevant framework for systems in our dataset, being systems in development or use in NSW. The NSW AIAF is the most detailed framework currently in place in Australia. It is also one of the longest-running regulatory frameworks targeting AI of compulsory application that we could identify in our research across Organisation for Economic Co-operation and Development ('OECD') jurisdictions. In short, these two frameworks represent appropriate and fruitful points of comparison for the present analysis, as they respond to similar objectives, and the regulatory clout of the EU may influence future developments of the NSW AIAF. For example, the authors are aware that, at the time of writing, the NSW AIAF is under review. Moreover, it seems likely that any review of the NSW AIAF will draw on international definitions and norms, such as the OECD principles, which in turn were influenced by a regulatory dialogue over the years with the EU,[30] and in parallel were key in regulatory developments in other liberal democracies such as the US.[31]

Certainly, the *EU AI Act* and the NSW AIAF are different in many respects. The *EU AI Act* is a supranational regulation, binding on Member States of the EU. As noted above, it seeks to regulate development, deployment, placement into market, and use of AI systems across both private and public sectors. In contrast, the NSW AIAF is an internal policy framework for NSW government agencies to apply when developing, deploying, or procuring AI technologies, to assess, in the first place, if that AI system poses a potential elevated risk according to guidance within the framework.[32] Whilst mandatory, it is not designed to be enforceable by third parties and can more readily be changed by the government at any time. Differences of both form and function between the *EU AI Act* and NSW AIAF flow through into differences in the drafting and framing of the rules and definitions. Generally speaking, the terms of the *EU AI Act* involve greater precision compared to the NSW AIAF. This is unsurprising considering the former is legally binding law of general application whilst the NSW framework is an internal policy mostly applied via self-assessment.[33]

---

30 Luca Bertuzzi, 'OECD Updates Definition of Artificial Intelligence "to Inform EU's AI Act"', *Euractiv* (online, 9 November 2023) <https://www.euractiv.com/section/tech/news/oecd-updates-definition-of-artificial-intelligence-to-inform-eus-ai-act/>.

31 José-Miguel Bello y Villarino et al, *Standardisation, Trust and Democratic Principles: The Global Race to Regulate Artificial Intelligence* (Report, July 2023) 8, 11 <https://www.ussc.edu.au/standardisation-trust-and-democratic-principles-the-global-race-to-regulate-artificial-intelligence>.

32 'NSW AIAF' (n 4) 6, 11. Further detail on when and how the NSW AIAF operates is provided below.

33 The NSW AIAF is primarily a self-assessment tool. If, after mitigation, a government department self-assesses their system to pose 'mid-range', 'high', or 'very high' residual risk, the system is considered an 'elevated' risk. Where the residual risk is 'high' or 'very high', the NSW AI Review Committee must review the self-assessment, and an independent risk audit is required. Where the risk is 'mid-range', an independent risk audit is required, and the department is told to consider review by the NSW AI Review Committee: ibid 65.

Notwithstanding these significant differences, both frameworks embody similar policy objectives and also approach the question of scope in similar ways. Both regulatory frameworks seek to balance two policy goals: (1) enabling the development and deployment of new digital technologies, including by public sector actors providing public services, and (2) ensuring that technological solutions are developed and deployed in a way that minimises risks of harm to the rights and interests of individuals,[34] and to other important societal interests.[35] To balance these goals, both frameworks adopt a risk-based approach. That is, in order to avoid imposing burdensome obligations on every data-driven technical system, the additional processes and standards set out in the regulations to identify and address or mitigate risks are intended to apply only to a subset of systems: those involving AI, and those which impose a higher or elevated risk.

A Contextualising the Analysis of the Two Frameworks

Our approach to analysing these two frameworks can broadly be described as a doctrinal comparative approach to the extent that we offer a doctrinal analysis of the application of these two frameworks and compare and contrast their respective approaches.[36] This enables us to ask questions about their interpretation and possible deficiencies, and provides the basis for normative questions about the proper scope of application for current and emerging legal frameworks for regulating new technologies in the public sector.

First, we briefly describe the operation of the two frameworks before closely analysing their respective approaches to defining AI, and to risk.

1 The Operation of the EU AI Act

The *EU AI Act* was passed in 2024. It is a legislative framework for AI systems commercialised or deployed across the public and private sectors of the EU and is envisaged as a Regulation with effects expanding to associated European Economic Area states (ie, Iceland, Liechtenstein, and Norway).[37] It prohibits some use cases for AI[38] and has separate rules for general purpose AI ('GPAI') and GPAI posing systemic risks,[39] but the most detailed regulatory provisions in the *Act* apply to high-risk uses of AI.

---

34 Various taxonomies have been developed to capture the different elements or dimensions of harm. For a recent systematic literature review of prior work, see Gavin Abercrombie et al, 'A Collaborative, Human-Centred Taxonomy of AI, Algorithmic, and Automation Harms' (2024) *arXiv* 2407.01294:1–21 <https://doi.org/10.48550/arXiv.2407.01294>. For a recent legal taxonomy of harms, see Sylvia Lu, 'Regulating Algorithmic Harms' (2025) *Florida Law Review* (forthcoming) <https://dx.doi.org/10.2139/ssrn.4949052>.

35 There are of course other goals and interests at play: the recitals of the *EU AI Act* (n 3), for example, outline many public policy aims and interests to be protected, and as supranational legislation a key goal of the *EU AI Act* (n 3) is to ensure a degree of consistency in regulation needed for the operation of the single market.

36 Theunis Roux, 'Comparative Public Law' (Paper, 2020) 5 <https://doi.org/10.2139/ssrn.3685535>.

37 *EU AI Act* (n 3) art 2(1).

38 Ibid ch II.

39 Ibid ch V.

Where a system is categorised as a high-risk AI system under the *EU AI Act* (see below Part III(C)(1)), developers must conduct thorough conformity assessments and establish continuous risk management to guarantee safety. They are required to use high-quality, unbiased data and ensure transparency through detailed documentation and traceability, alongside implementing some degree of human oversight and considering accuracy, robustness, and cybersecurity. They must also monitor the systems they develop after they have been placed on the market.[40] A different set of obligations applies to GPAI (not relevant here).[41]

2 The Operation of the NSW AIAF

The NSW AIAF was instituted in March 2022.[42] It is a mandatory self-assessment policy framework for NSW government agencies to apply when developing, deploying, or procuring AI technologies where that AI system poses a potential elevated risk according to guidance within the framework.[43] The framework does not apply where a government agency is using a widely available commercial application (which the agency is neither training nor customising) and it is also inapplicable where the agency is conducting exploratory research.[44]

Where a system is identified as posing potential elevated risk (examined in Part III(B)(2) below), the self-assessment framework requires those allocated as 'Responsible Officers' to work with team members to complete the self-assessment.[45] Responses must be recorded in the Record Management system.[46] The self-assessment requires the Responsible Officer to answer a series of questions relating to the NSW AIAF's five ethics principles for uses of AI: community benefit, fairness, privacy and security, transparency, and accountability.[47] The officer must summarise the risks identified under each principle and then record the mitigations to be applied to manage the identified risks.[48] If, after the consideration of all mitigations, the system remains at mid-range or higher residual risk, it is 'an Elevated risk use of AI'.[49] Different consequences flow from different final assessments of the extent of risk:[50]

---

40 Obligations in relation to high-risk systems are set out in chapter III of the *EU AI Act* (n 3).

41 *EU AI Act* (n 3) ch V.

42 'NSW AIAF' (n 4) 6.

43 'NSW AIAF' (n 4) 11, 13–14. Further detail on when and how the NSW AIAF operates is provided below. The NSW AIAF is an element of a broader Digital Assurance Framework ('DAF'), and non-AI projects may still be subject to various assessments under that DAF: Department of Customer Service (NSW), 'Digital Assurance Framework' (Framework, March 2024) <https://www.digital.nsw.gov.au/sites/default/files/2024-05/Digital-Assurance-Framework-2024.pdf>. The requirements of the DAF, however, are directed at more traditional Information Communications Technology project risks, such as complexity, risk of project failure, budget risk, etc.

44 'NSW AIAF' (n 4) 6.

45 Ibid 18.

46 Ibid 64.

47 Ibid 7.

48 Ibid 64.

49 Ibid 65.

50 Ibid 73.

- If after all mitigations any residual risk is *high or greater*, the agency must engage the NSW AI Review Committee. The Committee provides feedback and recommendations to improve the system, but the Agency Responsible Officers remain responsible for implementing mitigations, the impact and outcomes.[51] In other words, the Committee is an advisory body which does not have power to prohibit or direct changes to systems.
- If after all mitigations, any residual risk is *mid-range*, the agency must run a pilot before scaling up.
- If after all mitigations, any residual risk is *low* but there is potential for that risk to increase, agencies are advised to consider running a pilot.

Having outlined the basic operation of these two frameworks, we turn to a close analysis of how these frameworks define AI, beginning with the *EU AI Act*.

B Defining AI Systems

1 AI Systems in the EU AI Act

Article 3(1) of the *EU AI Act* defines an AI system. The definition is aligned with the OECD revised definition of AI,[52] which in turn finds its origins in the first OECD AI definition contained in the OECD Council's 2019 recommendation.[53] In other words, this definition, with some tweaks and adaptations, has been used in legal circles for a notable period – an age in AI time. It has seven elements: (1) a machine-based system; (2) that is designed to operate with varying levels of autonomy; (3) that may exhibit adaptiveness after deployment; (4) and that, for explicit or implicit objectives; (5) (5.1) infers, (5.2) from the input it receives, how to (6) generate outputs such as predictions, content, recommendations, or decisions (7) that can influence physical or virtual environments.[54]

Elements (1), (6), and (7) are applicable to most software systems. The traits which may differentiate AI from traditional human-written rules-based technical systems lie in the system's autonomy and adaptiveness (elements (2) and (3)); that its objectives may be 'implicit' (element (4)); and the fact that the system may 'infer' how to generate outputs from inputs (element (5)).

In understanding the various elements of the definition, we have the benefit of OECD commentary and guidelines published by the European Commission ('Commission') in February 2025 which offer authoritative, albeit non-binding,

51 Ibid 9.

52 Stuart Russell, Karine Perset and Marko Grobelnik, 'Updates to the OECD's Definition of an AI System Explained', *OECD.AI Policy Observatory* (Blog Post, 29 November 2023) <https://oecd.ai/en/wonk/ai-system-definition-update>.

53 Organisation for Economic Co-operation and Development, Council, 'Recommendation of the Council on Artificial Intelligence' (Council Minutes, No C/MIN(2019)3, 24 May 2019) 3 <https://one.oecd.org/document/C/MIN(2019)3/FINAL/en/pdf> ('Recommendation of the Council on Artificial Intelligence').

54 *EU AI Act* (n 3) art 3(1); European Commission, 'Commission Guidelines on the Definition of an Artificial Intelligence System Established by Regulation (EU) 2024/1689 (AI Act)' (Guidelines, No 5053, 29 July 2025) [9] ('Commission Guidelines').

interpretive commentary on what counts as an AI system under the *EU AI Act*.[55] These guidelines add interesting nuances to the definition, enabling us both to better understand each of the individual elements and to envisage the types of systems the Commission will seek to capture in future enforcement of the *EU AI Act*.

At the outset we note that the Commission considers that elements of the definition 'are not required to be present continuously throughout both phases of [the AI system's] lifecycle'.[56] In other words, a system that features elements from the definition (for example, autonomy) at some point or another of the lifecycle would qualify as an AI system for regulatory purposes, even if those elements are not present at the time of operation. This is also supported by one of the final observations in the guidelines stating that 'whether a software system is an AI system should be based on the specific architecture and functionality of a given system and *should take into consideration* the seven elements'.[57]

On objectives, the OECD's original definition of AI in 2019 referred to systems having 'a given set of human-defined objectives'.[58] In a short article on its website explaining the change to 'explicit or implicit' objectives, OECD staff noted 'implicit' could be understood in three different ways:[59]

1. *Objectives could be implicit in rules and policies*: a system given data and a broad objective (like 'be successful') could learn rules necessary for success and hence more specific objectives. For example, a system given the objective of winning chess games will learn the rules of chess if it is disqualified (and loses) every time it makes an illegal move. Here the rules of chess provide implicit additional objectives.
2. *Objectives could be implicit in training data*: the same system could learn that the objective of chess is winning and that only happens after checkmate if it sees enough games in the training data with the vast majority finishing that way. Here the final objective (winning) is implicit, as well as more specific objectives necessary to achieve that objective.
3. *Objectives may not be fully known in advance, but learned through interactions with humans*: for example, a film recommender system learns based on 'reinforcement learning from human feedback'. Systems will learn their objectives, based on observables such as how often humans interact with the system, given that, if the recommendations are useful, humans are more likely to use them.

55 These are developed in fulfilment of a mandate in article 96(1)(f) of the *EU AI Act* (n 3) to provide stakeholders with guidance about article 3(1): 'Commission Guidelines' (n 54) [3].

56 'Commission Guidelines' (n 54) [10].

57 Ibid [61] (emphasis added).

58 'Recommendation of the Council on Artificial Intelligence' (n 53) 3.

59 Marko Grobelnik, Karine Perset and Stuart Russell, 'What Is AI? Can You Make a Clear Distinction between AI and Non-AI Systems?', *OECD.AI Policy Observatory* (Blog Post, 6 March 2024) <https://oecd.ai/en/wonk/definition>.

Consistent with these comments from the OECD, the Commission has noted that the system objectives are not the same as the 'intended purpose' of the system, which is the external use intended by the provider of the system when deployed.[60]

Autonomy and adaptiveness are important elements of the definition. According to the guidelines, the Commission understands adaptiveness to mean 'self-learning capabilities, allowing the behaviour of the system to change while in use'.[61] As stated in article 3 of the *EU AI Act*, adaptiveness is a feature exhibited 'after deployment'. The drafting of the definition stating that a system 'may' exhibit adaptiveness after deployment suggests that adaptiveness is not an essential element to categorise a system as AI-based. The reverse may be true, however. A system having the capacity to adapt would likely qualify as an AI system. The guidelines support this interpretation.[62]

According to the *EU AI Act* definition, an AI system is 'designed to operate' with some degree of autonomy.[63] Here, the key factor would be the way 'some degree' is understood. Recital 12 of the *EU AI Act* clarifies that the expression 'varying levels of autonomy' means that AI systems are designed to operate with 'some degree of independence of actions from human involvement and of capabilities to operate without human intervention'. The guidelines note only that this 'excludes systems … designed to operate solely with full manual human involvement' but that there is a spectrum of 'human involvement' or 'human intervention' between fully human and fully automated.[64]

Importantly, however, the guidelines draw an explicit link between autonomy and inference, asserting that they 'go hand in hand: the inference capacity of an AI system … is key to bring about its autonomy'.[65] It is here that the Commission's guidelines impose some important limits on what counts as AI. The guidelines focus, both generally in the discussion and in outlining examples of systems that would *not* count as AI, on the ability of a system to infer autonomously or 'intelligently' and/or adapt its decision-making model.[66] The guidelines also explain what the Commission sees as current AI-related techniques – namely machine learning (including supervised learning, unsupervised learning, self-supervised learning, reinforcement learning, and deep learning), and logic- and knowledge-based

60 'Commission Guidelines' (n 54) [24]–[25].
61 *EU AI Act* (n 3) recital 12.
62 'Commission Guidelines' (n 54) [23].
63 *EU AI Act* (n 3) art 3(1).
64 'Commission Guidelines' (n 54) [16]–[17].
65 Ibid [15].
66 Ibid [45]. Interestingly, the Commission notes that its conceptualisation of inference is compatible with the definition included in the ISO/IEC 22989 standard ('reasoning by which conclusions are derived from known premises'): International Organization for Standardization, 'ISO/IEC 22989:2022(en)', *Online Browsing Platform* (Web Page, 2022) cl 3.1.17 <https://www.iso.org/obp/ui/#iso:std:iso-iec:22989:ed-1:v1:en>. See the discussion of that standard in José-Miguel Bello y Villarino, 'Global Standard-Setting for Artificial Intelligence: Para-regulating International Law for AI?' (2023) 41 *Australian Yearbook of International Law* 157 <http://dx.doi.org/10.1163/26660229-04101018>. However, by referring to specific technical techniques, the Commission introduces a much narrower understanding of the type of 'reasoning' which counts, as compared to the ISO standard.

approaches[67] – and what it does not: namely 'basic data processing' or 'systems [that] have the capacity to infer in a narrow manner'.[68] It seems, both from this text and from more detailed examples discussed in the guidelines, that the Commission sees the *EU AI Act* as applying to systems that can show a level of 'intelligence' to infer outputs, beyond the traditional techniques like regression, and in a way that may be difficult to explain. In other words, the Commission seems to target 'black box' systems.[69]

In summary: according to current guidance from the Commission, AI systems are defined by the presence of autonomous or intelligent inference and/or intelligent adaptiveness, with certain specific technical methods (supervised learning, unsupervised learning, self-supervised learning, reinforcement learning, deep learning and logic- and knowledge-based approaches) seen as clear and core present examples of AI. We apply this definition in our empirical analysis below.

2 AI in the NSW AIAF

Compared with the *EU AI Act*, the NSW AIAF adopts far looser language in talking about what constitutes AI. Recall our description of the NSW AIAF above. The NSW AIAF was instituted in March 2022 with the stated goal of '[guiding] NSW government agencies through the ethical development, deployment and use of artificial intelligence (AI) technologies'.[70] It is a mandatory framework for NSW government agencies to apply when developing, deploying, or procuring an AI system where that system poses a potentially elevated risk according to guidance within the framework.[71] A revised version (renamed from the AI *Assurance* to the AI *Assessment* Framework) was published in July 2024. At the time of writing, the NSW AIAF is set out in two separate documents published online:[72] a Microsoft Excel table, designed for use in the assessment process, and a more reader-friendly Microsoft PowerPoint presentation.

The official NSW government website where the NSW AIAF is published acknowledges that there is no 'universally accepted technical or legal definition of AI', and offers only a high-level definition of AI as 'the ability of a computer system to perform tasks that would normally require human intelligence, such as

67 'Commission Guidelines' (n 54) [30]–[39].

68 Ibid [41]–[42]. Here, the guidelines take an interesting detour into the past. In the first version of the proposal for an *EU AI Act*, the Commission included this same list of technologies and techniques as ones that would 'characterise' AI systems: *EU AI Act (Proposal)* (n 26) art 3(1), annex I. The Annex that contained that list has now disappeared from the binding part of the regulation, but subsists in recital 12, which directs stakeholders to a series of techniques that enable systems to infer how to generate outputs. The Commission in its guidelines develops recital 12: ibid [40]–[51].

69 Georgios Pavlidis, 'Unlocking the Black Box: Analysing the EU Artificial Intelligence Act's Framework for Explainability in AI' (2024) 16(1) *Law, Innovation and Technology* 293 <https://doi.org/10.1080/17579961.2024.2313795>.

70 'NSW AIAF' (n 4) 6; NSW Government, 'NSW Artificial Intelligence Assessment Framework', *Digital NSW* (Web Page) <https://www.digital.nsw.gov.au/policy/artificial-intelligence/nsw-artificial-intelligence-assessment-framework> ('NSW AIAF Web Page').

71 'NSW AIAF' (n 4) 11.

72 'NSW AIAF Web Page' (n 70).

learning, reasoning, and making decisions'.[73] It also offers some guidance on the interpretation of this very open-textured definition on the website.[74] That guidance too uses open-ended language and specifically acknowledges the lack of any clear definition. As is the case in the EU, the current NSW guidance suggests that 'simple' systems should be excluded.

As we noted above, the lack of clarity in the NSW AIAF regarding what is intended to count as AI is possible in part because of the function of the framework as a mandatory but still internal tool based on self-assessment. The *EU AI Act*, as a law seeking to bind actors across multiple jurisdictions and requiring cooperation from multiple national governments and domestic regulators, could not afford to be so open-ended: imposition of legal compliance calls for greater bright line certainty. Consistent too with its function, guidance provided by the NSW Government on the NSW AIAF suggests that the self-assessment should be undertaken where there is a *possibility* of AI use – ie, when in doubt, public servants are advised to apply the framework.[75]

In the particular context of our empirical analysis, we have found that it is impossible to use or apply the NSW AIAF understanding of what constitutes AI. We are not public servants with full information about the systems in our dataset. We must base our assessment on brief descriptions of the systems and some minimal information about the technologies used. We have therefore classified systems as 'AI or not' based exclusively on the *EU AI Act* definition and associated Commission guidelines. Not only is the *EU AI Act* definition aligned with a widely accepted international definition provided by the OECD, but it is also influential in Australian legal and policy circles. The same definition is used in various Australian policy documents, including the Commonwealth Government's proposals paper for introducing mandatory guardrails for AI in high-risk settings.[76] We also consider it fair to assume that any system identified as AI on that basis would also be classified as AI by users of the NSW AIAF. We note, however, that the reverse is not true. The open-ended definition, and 'when in doubt include' interpretive approach suggested in the NSW AIAF could mean some systems are and/or ought to be assessed under the framework which would be excluded from the regulatory requirements of the *EU AI Act*. On the other hand, it is unclear how far an expansive approach is likely to be applied in practice. Public servants will have disincentives to adopt an expansive view: classifying a system as involving AI leads to additional work, assessment, risk mitigation requirements and possible external scrutiny.

---

73 NSW Government, 'Identifying AI', *Digital NSW* (Web Page, 19 December 2024) <https://www.digital.nsw.gov.au/policy/artificial-intelligence/nsw-artificial-intelligence-assessment-framework/identifying-ai> ('Identifying AI'). The definition of AI as, broadly, 'machines designed to do things that require human intelligence' is a classic understanding of the goal of the field of AI, dating from the 1950s: see generally Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach* (Pearson, 4th ed, 2022) 19–23.

74 'Identifying AI' (n 73).

75 See, eg, 'NSW AIAF' (n 4) 11.

76 'Safe and Responsible AI in Australia' (n 29) 8.

C Classifying Systems as 'High' or 'Elevated' Risk

The second key factor that determines the scope of both the *EU AI Act* and the NSW AIAF is whether a system in use is considered to pose a sufficiently serious risk of harm to individual and societal interests to require further consideration and/or governance in order to avoid or mitigate those risks. The *EU AI Act* describes within-scope systems as 'high-risk' systems. The NSW AIAF uses slightly different language, suggesting that the assessment framework should be used where a system poses 'potential elevated risk'. In order to test these regulatory frameworks against our dataset, then, we need to understand what constitutes a system with 'high-risk' (in the EU) or 'potential elevated risk' (in NSW). As we explain here, the two regulatory frameworks differ significantly in the way they classify systems. The *EU AI Act* classifies high-risk systems based on use cases and domains of deployment; the NSW AIAF, by contrast, focuses on a range of indicators of risk attached to the system and its context of use.

1 High-Risk Systems in the EU AI Act

Under article 6(1) of the *EU AI Act*, an AI system will be classified as a high-risk system if it falls into one of two categories:

(a) If it is a safety component of a product or is itself a product covered by one of the pieces of 'harmonisation' product safety legislation listed in Annex I of the *Act*. Examples of the legislation listed in Annex I include product safety directives relating to toys, machinery, lifts, personal protective equipment, and medical devices.

(b) If it is to be used for one of the applications set out in Annex III of the *Act*.

Annex III sets out a list of specific applications that fall within the following general categories:

- biometric identification and categorisation, including emotion recognition;
- management and operation of critical infrastructure;
- certain systems within educational and vocational training (such as systems determining access and admission, evaluating learning outcomes, and systems for monitoring students during tests);
- employment (certain systems within hiring), worker management, and access to self-employment;
- access to and enjoyment of essential private and public sector services and benefits;
- certain use cases within law enforcement;
- migration, asylum, and border management; and
- administration of justice and democracy.

Note that we have left out details here: any AI system used in the listed areas of activity is deemed high-risk under Annex III. In other words, the *EU AI Act* defines high-risk systems by reference to certain narrow use cases. Article 6 further provides a derogation for systems that fall within the Annex III list but do 'not pose a significant risk of harm to the health, safety, or fundamental rights of natural persons, including by not materially influencing the outcome of decision making'.

Under article 6(3), this is said to be the case if a system that is not used to profile natural persons is intended to:

(a) perform a narrow procedural task; or
(b) improve the result of a previously completed human activity; or
(c) detect decision-making patterns or deviations from prior decision-making patterns, not replacing or influencing the previously completed human assessment without proper human review; or
(d) perform a preparatory task to an assessment relevant for the purposes of the high-risk use cases listed in Annex III.

Article 6(5) also requires the Commission to provide guidelines specifying the practical implementation of the article, together with a comprehensive list of practical examples of use cases of AI systems that are high-risk and not high-risk, by 2 February 2026. Even without further guidelines, however, the existing *EU AI Act* text provides us with a relatively clear definition which we apply below to our dataset in order to determine whether systems in development or use as of early 2024 in NSW would be considered high-risk within the EU.

2 Elevated Risk under the NSW AIAF

The NSW AIAF requires 'all NSW Government Agencies'[77] intending to buy and use AI, embed and/or co-train AI, develop and/or train AI, or automate decisions with a tool that uses AI to first undertake a determination of whether the AI system gives rise to potential elevated risk.[78] If a system meets the threshold of potential elevated risk, and involves AI in one of the above-described ways,[79] then the system falls within scope and the agency is required to apply the self-assessment framework and review the summary risk.

All self-assessments, explanations, and mitigations must be recorded in the Record Management system.[80] If after conducting the self-assessment and 'considering all mitigations' the system remains at 'mid-range' or 'high' residual risk there are additional obligations imposed.[81] Where after applying the risk mitigations the residual risk remains 'high', the system must be referred to the NSW AI Review Committee.[82] Where the residual risk is assessed at 'mid-range', a pilot of the system must first be undertaken.[83]

The framework provides six independent indicators for when an AI system poses potential elevated risk. Only one of the indicators needs to be present. As with the question, whether a system involves AI, the framework notes that public servants will need to make a judgment about potential elevated risk, based on the particular use case. If they are unsure, the framework instructs public servants

77 'NSW AIAF' (n 4) 6.
78 Ibid 11, 13.
79 Ibid 13.
80 Ibid 64.
81 See ibid 65.
82 Ibid 73.
83 Ibid.

to assume the system *does* pose potentially elevated risk.[84] The indicators are expressed in terms of questions agency staff should ask themselves:[85]

- *Operational impact*: Does your system produce or directly influence an administrative decision (government decisions with legal or similar significant effect)?
- *Operational impact*: Does your system trigger a real-world action with more than negligible potential effect (meaningful change to environment or system state)?
- *Autonomous*: Does your system operate autonomously or have potential to produce harmful outputs independently of human action, without requiring manual initiation?
- *Data sensitivity*: Was any part of the system trained using sensitive information or can it produce outputs which contain sensitive information?[86]
- *Unintended harms*: Is there a risk of system failure, misuse, or inappropriate deployment that could cause harm to an individual or group?
- *Explainability and transparency*: Does your system fail to provide explainability for generated content and decisions, hindering comprehension by laypeople and assessment by technical experts?

The NSW AIAF also provides examples for each indicator (see below Figure 1).[87]

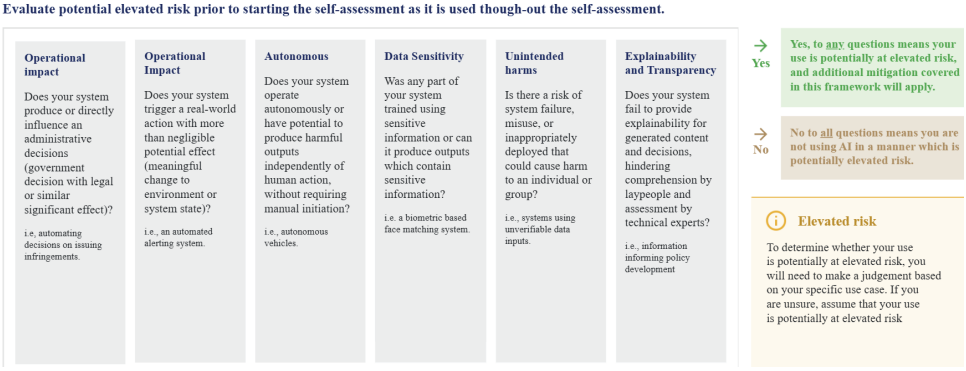

Figure 1: Indicators by area for use cases of potential elevated risk in the NSW AIAF.

Answering 'no' to all of these questions removes the use case from the scope of the framework, even if other procedural and substantive obligations in law and policies would still apply (eg, cybersecurity or fairness considerations).

84 Ibid 11.

85 Ibid.

86 In NSW, sensitive information includes an individual's personal details, credit information, medical records, driver licence information, criminal records, and biometric information: see, 'NSW Government Information Classification, Labelling and Handling Guidelines' Sensitive Information' (Guidelines, No 2.3, Customer Service, April 2021) 15 <https://data.nsw.gov.au/data-policy/nsw-government-information-classification-labelling-and-handling-guidelines>.

87 'NSW AIAF' (n 4) 11.

IV EMPIRICAL ANALYSIS: TESTING THE FRAMEWORKS AGAINST A REAL-WORLD DATASET OF PUBLIC SECTOR ADM SYSTEMS

So far, we have discussed the two criteria for technical systems to fall within (a) the *EU AI Act* and/or (b) the NSW AIAF. Those two criteria are (1) the system involves the use of AI, and (2) the AI system reaches a threshold of 'high' or 'potential elevated' risk. Applying these criteria to a real-world list of automated systems enables some interesting analysis: in particular we can 'pick apart' the two criteria, and consider which systems involve high-risk uses of technology but do not involve AI. The implications and relevance of disaggregating the classification of systems as AI and 'high-risk', in particular if we find that many systems fall into the category of non-AI high-risk, will be considered in the discussion below.

A The Dataset

The dataset we analyse in this article – which we also make accessible to readers – derives from data about ADM systems in development or use in the NSW government from two sources. The main source is a dataset that was developed as part of a research project commissioned by the NSWOO to map the utilisation of automated systems in administrative decision-making processes within NSW state departments and agencies and local governments (the 'Mapping Project').[88] This research was the first attempt to undertake such a systematic mapping in any jurisdiction across Australia. Globally, this initiative stands as one of very few comprehensive attempts to create a systematic mapping of automation in the public sector.[89] The methodology of the Mapping Project, including full copies of the research instruments, is set out in appendices to the final full report of the Mapping Project.[90] To provide context to the reader, we offer a summary of key elements relevant to the analysis below.

For the purposes of the current analysis, we incorporated data collected during the first stage of the Mapping Project through a short survey sent to every department and agency of the NSW state government and every NSW local council. The first survey asked each department, agency, and local council to identify every ADM system planned, in development, in pilot, in use, or previously used but no

88 Weatherall et al (n 5).

89 The Public Law Project in the UK has a similar project called the Tracking Automated Government ('TAG') Project which has produced and updated a record of automated tools known as the 'TAG Register'. It currently records 55 automated tools operating in the UK. See 'The Tracking Automated Government Register: A New Tool Lifting the Lid on Secret Government Algorithms', *Public Law Project* (Web Page, 9 February 2023) <https://publiclawproject.org.uk/resources/the-tracking-automated-government-register/>. The NSW project is unique, however, in having the cooperation of a public agency, which may have increased the number of systems identified. We also note that a number of governments internationally are creating algorithmic transparency registers of various kinds, such as the UK's algorithmic transparency records: Cabinet Office (UK), Department for Science, Innovation and Technology (UK) and Government Digital Service (UK), 'Find Out How Algorithmic Tools Are Used in Public Organisations', *GOV.UK* (Web Page) <https://www.gov.uk/algorithmic-transparency-records>.

90 Weatherall et al (n 5) 121–38. Research instruments are included in the Appendix: at 139–82.

longer being used within their organisation.[91] A second survey was subsequently conducted to obtain more detailed information about a smaller subset (n = 34) of ADM systems identified in the initial survey.[92]

The Mapping Project sought to identify ADM systems, which were defined for the purposes of the project as fully or partially automated technical systems used by a NSW government organisation (state government department or agency, or local council) in administrative decision-making that affected members of the public.[93] The survey instruments specifically noted that the researchers were interested in:

> ADM systems that automate, whether fully, partially or as part of a bigger decision-making process, the exercise of government functions, and in doing so impact individuals or private entities. We are also interested in systems that members of the public interact with in relation to government activities. We are interested in both in systems [sic] that use artificial intelligence/machine learning, and those which use traditional forms of computer programming.[94]

The second source for the consolidated dataset used in our analysis is data directly collected by the NSWOO in January–February 2024 to confirm and expand where necessary the dataset collected during the Mapping Project. This second source dataset is publicly accessible: it was published by the NSWOO as the *Compendium of ADM Systems* ('*Compendium*') in March 2024.[95] For the purposes of this article, the NSWOO provided some further details about the technologies used for each system able to be identified by a specific name and listed in the *Compendium*. Nevertheless, for systems identified only in the second dataset, there may be more limited information than in the first dataset because the first dataset includes answers to additional questions posed in the first survey.

Combining data from these two sources created a consolidated dataset of 266 ADM systems identified by name and accompanied by short textual descriptions. This list was further refined. First, we excluded all systems deployed by local government, leaving only systems used by NSW state departments and agencies in the dataset. Second, we only included systems that were identified by the responsible entities as 'in use' or 'in development', which excluded systems listed in the *Compendium* but only planned for possible development over the next three years or those recently discontinued. This left us with a list of 163 systems which, at the time of the research, were either firmly 'in development' or 'in use' (comprising approximately 61% of the total ADM systems identified). These ADM systems are spread across all NSW state government portfolios and multiple departments and agencies, reflecting the diverse applications and purposes of ADM systems in the NSW government. It should be noted here that, despite the best efforts of the

---

91 Ibid 124–5. For systems 'planned' the survey asked only for those planned within the next three years. For systems 'previously used', the survey similarly asked for systems that had been in use within the previous three years.

92 Ibid 129–30. The survey process was led from the University of Queensland by Professor Paul Henman, Chief Investigator at ADM+S and included extensive work to ensure a maximum number of departments and agencies were reached as well as follow up processes: see ibid 126–9.

93 Ibid 124.

94 Ibid 159–60.

95 *Compendium of ADM Systems* (n 6).

research team, the data collected did not include any systems used by the NSW Police, which is outside the NSWOO mandate.

A table of the ADM systems is publicly available for the reader.[96]

B Methodology for Classifying Systems in the Dataset

Taking the 163 systems as the total universe for our research (n = 163), we then used the text descriptions to determine whether each system could be considered to give rise to potential elevated risk within the NSW AIAF. In parallel, we considered whether each system could potentially be high-risk within the meaning of the *EU AI Act*. We performed this analysis based on the use cases as publicly listed and by reference to the domains where the systems are already deployed or are expected to be deployed. This was followed by an assessment of whether each system would likely be classified as 'AI or not' under the *EU AI Act* definition.

For the assessment as high risk under the *EU AI Act* we covered the three possible bases for such a classification provided for in the Act; that is, whether the use case fits one of the categories described above in Part III(C). For the NSW AIAF, we sought to determine the presence of any one of the six indicators of potentially elevated risk.

As foreshadowed, the broad, open-ended description of AI in the NSW AIAF does not readily lend itself to the classification of systems we sought to conduct: ie, on the basis of short text descriptions of each system. More details, such as the information which would normally be in the hands of a developing or deploying government department or agency would be required. We therefore classified systems as 'AI or not' only by reference to the *EU AI Act* definition discussed above in Part III(B)(1).

The likelihood that any one system would meet the respective thresholds for (i) risk and (ii) AI assessment, according to the noted criteria, was first given a rating along a spectrum: 'Yes'/'Probably'/'Possible'/'Unlikely'/'No'/'Insufficient information'. For purposes of the further analysis, following that initial classification, we consolidated the ratings into three options: 'Yes' ('Yes'/'Probably'); 'No' ('No'/'Unlikely'); and 'could not tell' ('Possible'/'Insufficient information').

In order to ensure consistency in classification, these evaluations were conducted by one member of the research team across the three parameters. That researcher conducted three separate and independent classification exercises: ie, the evaluation was done for one parameter at a time without seeing or considering how the systems had been previously classified according to the other parameters. To mitigate the risk that memory retention of previous evaluations could introduce some unconscious bias, these evaluation exercises were spaced out over several weeks. The alternative of multiplying the number of evaluators was discarded, as it would have required training several people on the nuances of the definitions and categories established by the NSW AIAF and the *EU AI Act*. We are also

96 We have made the dataset available here: José-Miguel Bello Villarino and Alexandra Sinclair, 'Analysis of Risk Levels and Use of AI in Automated Decision-Making Systems in NSW', *University of Sydney* (Repository) <https://doi.org/10.25910/7ccd-nx49>.

making the full dataset available, meaning all our classifications are available for independent review.

To increase the methodological reliability for the most sensitive part of the analysis – the classification of systems to be of 'potentially elevated risk' in the NSW AIAF – we ran a parallel quality check by a second evaluator, who categorised a randomly chosen sample of one fifth of the 163 systems (n = 34). Their analysis showed a reasonable degree of consistency in three-way options – 'Yes' ('Yes'/'Probably'); 'No' ('No'/'Unlikely'); 'could not tell' ('Possible'/'Insufficient information') – at over 50% (n = 18; 53%), against a random chance of alignment of 1 in 3 (33%). The 16 cases of misalignment of the assessment could be classified into five groups.

- **Group 1** comprised nine cases where the divergence was between the primary evaluator noting 'Yes' or 'No' and the secondary evaluator noting that they 'could not tell'. Further exploration suggested this was probably explained by the background knowledge from the main evaluator about some of the systems beyond the literal description.
- **Group 2** comprised two cases where the situation was the opposite.[97] The main evaluator 'could not tell' ('Possible'/'Insufficient information') and the secondary evaluator considered that they were not of 'potentially elevated risk' in the NSW AIAF (ie, rating 'No').

For the purposes of this article, neither of these two groups of divergent assessments (11 of 16) is particularly problematic. In those cases, the problem seems to be attached to the level of detail of the dataset or linked to the human background knowledge of the information in the datasets. In a real assessment setting, requesting further information would be expected to resolve this divergence. Furthermore, for Group 2, the opinion of the second evaluator was that these systems should not be classified as being of potentially elevated risk. Therefore, the risk of harm in a real setting derived from a misclassification in those cases seems of lesser relevance.

The other instances (5 of 16) of disagreeing evaluations require deeper exploration. There were three instances where the main evaluator said 'No' ('No'/'Unlikely') and the secondary evaluator classified it as a 'Yes' ('Yes'/'Probably') (Group 3); one (1) where the primary evaluator 'could not tell' and the secondary evaluator noted a 'Yes' (Group 4); and one (1) case where the main evaluator considered it of 'potentially elevated risk' and the secondary evaluator did not (Group 5).

- **Group 3:** The three instances of 'No' (primary) – 'Yes' (secondary) disagreement were chatbots.[98] For these cases, the second evaluator considered that, given their autonomy, these systems should qualify as 'potentially elevated risk' in the NSW AIAF, whereas the main evaluator considered that the risk of the autonomy was below the threshold of the NSW AIAF given

97 Eg, DeliverEASE (Scheduled Ordering and Security Information and Event Management (SIEM)) and Security Orchestration, Automation, Response (SOAR).

98 Visitor Itinerary Planner, Einstein Chatbot, Virtual Contact Centre – Digital Chat Bot.

the narrow scope of their respective outputs (tourist/visitor itineraries, replies about leave, and replies within a given webpage content).

This disagreement may point to the complexity of assessing the risk of chatbots *ex ante*. Chatbots in the *EU AI Act* are subject to a distinct regulatory regime. In addition to being assessed like all AI systems for whether they are high-risk or not depending on the use case, an additional transparency requirement applies. Individuals interacting with a chatbot must be informed that they are interacting with an automated system.[99] Nevertheless, for the purpose of our analysis, this divergence is of more limited relevance. The classification of these chatbots may slightly distort the data presented below, but the actual implications in terms of risk are extremely limited.

Two final instances of divergence in assessment may have had further implications.

- **Group 4:** NSWTG OWLS is described as an online wills system which makes risk-based decisions on instruments needing to be examined by a senior staff member.[100] The main evaluator considered that it was possible that it was of 'potential elevated risk', but 'could not tell'. The secondary evaluator considered there was 'potential elevated risk', as it had operational impact influencing an administrative decision. The difference between evaluators seemed to be on the interpretation of when a system will 'produce or *directly influence an administrative* [*decision*] (government decision with legal or similar significant effect)' in the NSW AIAF.[101] The logic of the main evaluator was that it was unlikely (although possible) that it would *directly* influence an administrative decision, which remained fully with the senior staff member. This type of divergence could also manifest in the *EU AI Act* context in relation to article 6(3)(d), which exempts from high-risk systems those that 'perform a preparatory task to an assessment'.
- **Group 5:** The Toll Relief Rebate system is described as 'a transaction in which motorists can claim back parts of their toll spend. Eligibility is determined based on the eligibility criteria set out in the policy.'[102] The system triggers an automated payment, based on records of past use of the tolled motorways. The main evaluator considered the Toll Relief Rebate system of potential elevated risk and the secondary evaluator did not. The main evaluator considered the system produced an administrative decision with legal or similar significant effect. Given the direct effect and the scope of the population affected, it seemed that the NSW AIAF would intend to capture these use cases.

Despite these two instances, in general terms the quality check confirmed the reliability and above-minimal validity of the evaluation of the classification of potentially elevated risk under the NSW AIAF. As this parameter was the keystone

99 *EU AI Act* (n 3) art 50(1).
100 *Compendium of ADM Systems* (n 6) 8.
101 'NSW AIAF' (n 4) 11 (emphasis added).
102 *Compendium of ADM Systems* (n 6) 15.

of the broader analysis, we proceeded to explore the relationship between the three chosen parameters, ie, high-risk under the *EU AI Act*, potentially elevated risk under the NSW AIAF, and whether the system is AI according to the *EU AI Act*.

C Analysis: What Do the Numbers Tell Us About the Scope of the *EU AI Act* and NSW AIAF?

1 High-Risk (EU AI Act) vs Elevated Risk (NSW AIAF)

We were able to provide positive risk ratings of 'Yes' ('Yes'/'Probably') or 'No' ('No'/'Unlikely') for just under two-thirds of our 163 systems (n = 104; 64%). The remaining systems (n = 59; 36%) received a rating of 'could not tell' as we either did not have sufficient information or could not lean clearly towards a 'Yes' or 'No', instead considering them only 'Possible'. As expected, there is a significant overlap between systems that could not be clearly categorised as high- or low-risk under the *EU AI Act* and the NSW AIAF.

For the 104 systems with meaningful classifications (simplified to 'Yes'/'No'[103]) we conducted a descriptive concordance analysis in pairs. The results are condensed in Table 1.

Table 1: Concordance analysis of dataset against risk thresholds

Concordance Analysis		Potentially Elevated Risk in the NSW AIAF	
		No	Yes
High-Risk in the *EU AI Act*	No	51	18
	Yes	10	25

As shown in Table 1, there is broad consistency in the technical systems which, in our evaluation, both regulatory frameworks classify as posing (or not) potentially high (or elevated) risk. There are 51 instances where both columns have the value 'No' and 25 instances where both columns have the value 'Yes.' In other words, close to three-quarters of this subset of the systems are classified in the same way in both columns (76 of 104; 73%).

This may be slightly counterintuitive given the very different ways these systems approach the question of risk, as discussed in Part III(C) above. In our view, however, the significant overlap is not only comforting, it is explicable. Our data analysis here is only considering technical systems used in the public sector. This is one of the main differences between both regulations, as the EU also covers AI uses in the private sector.[104] Within the public sector, the overarching considerations of risk (potential harms) can reasonably be assumed to be similar across liberal-democratic jurisdictions. We would therefore anticipate that, regardless of how the

103 These simplified risk ratings are found in column 3 ('Simp_HighREU') for the *EU AI Act* simplified rating and column 5 ('Simp_ERNSW') for the NSW AIAF simplified rating.

104 *EU AI Act* (n 3) art 2.

regulatory scope is defined, the type of systems targeted should be similar. This suggests that the systems assessed as low-risk in the EU are often also assessed as low-risk in NSW, and similarly, those assessed as high-risk in the EU are likely to be often assessed as high-risk in NSW.

The more interesting space for explanation is where we find *discrepancies*: where we have classified a system under one framework as potentially of higher risk but reached the opposite conclusion under the other framework. Generally, the data suggests that the NSW AIAF captures more systems 'of relevant risk' in the public sector (18 + 25 = 43, Table 1) than does the *EU AI Act* (10 + 25 = 35, Table 1). This could be intentional, reflecting the different functions of these two regulatory frameworks. That is, a higher threshold for identifying a system as high-risk could be appropriate for the *EU AI Act*, where the rating triggers a more stringent set of regulatory requirements. Recall too that the NSW AIAF is framed in a more expansive way: it requires actors to identify *potential*, not *actual* elevated risk, and public servant users of the framework are urged to include systems in cases of doubt.

On the other hand, for the roughly one quarter (10 + 18 = 28, Table 1) of the relevant instances where the values differ, there is a relatively balanced distribution among them. In 18 instances (17%), the evaluation considered that it would be a system of potentially elevated risk in NSW but not high-risk in the EU, and for 10 instances (9.6%) the evaluation classified the systems in an inverse manner ('Yes' high-risk for the EU and 'No' potentially elevated risk under the NSW AIAF). It is then not possible to posit a direct correlation or link between the higher total number of instances of elevated risk classification under the NSW AIAF to a more expansive regulatory scope. For that explanation to hold, it would be necessary to find a more significant divergence between those quadrants, with lower numbers in the bottom left quadrant.

2 What Proportion of Potentially Elevated Risk Systems under the NSW AIAF Involve AI?

A key question of interest for policymaking in this area, and a core concern of this article, is the question of whether regulatory frameworks applicable only to AI systems capture all the systems deserving of more scrutiny and governance. We therefore also wanted to analyse whether systems considered to pose potentially elevated risk under the NSW AIAF would be considered to involve AI. As discussed above in Part III(B)(2), systems which do not involve AI would not be assessed under the framework, no matter how impactful they might be or what the risks are of unintended harms. As also noted above in Part IV(B), we did not find it possible to apply the NSW AIAF guidance on what constitutes AI.

Instead, we use here the *EU AI Act* definition of AI, which is based on the widely used OECD definition.[105] It is worth repeating here a qualification we noted above. While we are confident that systems classified as AI under the *EU AI Act* would almost certainly be treated as AI under the NSW AIAF, the reverse is not

105 See Bertuzzi (n 30).

necessarily true in all cases, owing to the open-ended text and 'when in doubt include' approach suggested in the NSW AIAF. We acknowledge that in adopting the *EU AI Act* definition for classification purposes, we risk being under-inclusive, as compared to the intended application of the NSW AIAF.

In classifying our systems as 'AI or not', we followed the same process as we used to classify them as high or elevated risk, described above. First, we determined for which of our total population of systems in development or use across NSW state government departments and agencies (n = 163) we had sufficient data to undertake the required analysis on *both* questions: whether the system involved AI, and whether it would be classified as posing potentially elevated risk under the NSW AIAF. Of the initial set of 163 systems, our evaluator was able to classify 113 (69%) systems on both variables with reasonable confidence (that is, a classification according to our simplified ratings of 'Yes' ('Yes'/'Probably') or 'No' ('No'/'Unlikely') answer for both variables). Of those 113 systems, we categorised just over half (n = 64; 57%) as not considered to be of potentially elevated risk in NSW and just under half (n = 49; 43%) as systems of potentially elevated risk. We set out the data below in Table 2.

Table 2: Concordance analysis of systems for risk (NSW AIAF) v AI (*EU AI Act*)

Concordance Analysis		AI System in the *EU AI Act*	
		No	Yes
Potentially Elevated Risk in the NSW AIAF	No	34	30
	Yes	29	20

As set out in Table 2, out of the 64 systems *not* considered to pose 'potentially elevated risk' in the NSW AIAF, 34 (53%) would not satisfy the *EU AI Act* definition of AI, while 30 (47%) did fall within the *EU AI Act* definition: a roughly even distribution. Of the 49 systems that *were* considered to pose 'potentially elevated risk', our evaluator classified 29 (59%) as not involving AI under the *EU AI Act*, while 20 of the 49 systems of potentially elevated risk (41%) were likely to qualify as AI systems in the EU.

For regulatory and policy purposes, the key quadrant in Table 2 is the bottom right one. Of the 49 systems we identified as posing potentially elevated risk, only 20 – a minority – would be subject to mandatory assessment and mitigation of risk under the NSW AIAF. For these 20 systems, the relevant agencies would be required to conduct a proper assurance assessment, with the possibility of being further mandated to submit their completed AI self-assessment to the AI Review Committee 'if the residual risk remains high or greater following the application of all mitigation and controls'.[106]

106 'NSW AIAF Web Page' (n 70). See above Part III(B)(2). Note that this only applies insofar as the system also meets the other scope-related requirements in the NSW AIAF in terms of the system not being a 'widely available commercial application' which is not being trained or customised: 'NSW AIAF' (n 4) 6.

However, these 20 systems are not our main concern here. If the goal of the NSW AIAF is to reduce harms arising from the deployment of technical systems in the public sector, it is the bottom left quadrant which is truly problematic from a regulatory perspective. As noted, 29 of the 49 systems that are considered to carry a potentially equivalent level of risk – or roughly 60% of the potentially elevated risk systems – are nonetheless exempted from any *ex ante* assessment or risk mitigation under the NSW AIAF.

A review of the systems falling into this category shows that more than 20 systems in the dataset, including some used in the most potentially controversial domains, were described using terms such as 'rule-based', 'sequential rules', 'based on ruleset', 'utilising business rules', or 'automated decision trees'. Consistent with the *EU AI Act* definition, we classified such systems as not involving AI. According to the Commission, a rules-based system will not be considered an AI system.[107]

Examples of non-AI, rules-based systems that operate in sensitive areas where care, scrutiny, and governance might be expected include:

- 'Rippledown', a system used by NSW Health Pathology which '[processes] clinical orders and billing to identify anomalies based on a set of sequential rules';[108]
- 'OIMS', used by Corrective Services NSW, which is '[a] partially automated decision tree with manual functions to determine the classification of inmates';[109]
- '[System 4]', a structured decision-making tool used by the Department of Communities and Justice (Courts, Tribunals and Service Delivery) to '[onboard] and manage defendants to the NSW Drug Court program';[110]
- '[System 5]', a structured decision-making tool used by the Department of Communities and Justice (Courts, Tribunals and Service Delivery) which is a 'Client Management System for victims of crime';[111]
- The 'online birth registration system', a fully automated rule-based system used by the Registry of Births, Deaths & Marriages 'utilising business rules to assist in confirming birth registration submissions from parents … [if] all rules are met birth can be registered automatically';[112]
- 'Byte', a fully automated rule-based system used by the NSW Food Authority for the 'bulk processing of licenses [sic] for various stages of the License lifecycle, based on coded business rules'.[113] It also uses Byte mobile, a fully automated rule-based system 'to determine audit and inspection outcomes based on level of compliance with inbuilt checklists'.[114]

107 'Commission Guidelines' (n 54) [40].
108 *Compendium of ADM Systems* (n 6) 26.
109 Ibid 7.
110 Ibid.
111 Ibid.
112 Ibid 12.
113 Ibid 29.
114 Ibid.

We emphasise that our analysis says nothing about how these systems listed immediately above were designed or developed; for all we know, each of these systems have been developed and deployed and are monitored with great care. The fact that a system falls outside the NSW AIAF says nothing about any risk assessment processes that might have been followed, although we suspect that the analysis of those Information and Communications Technology ('ICT') projects would be focused on what we could call 'traditional' ICT risks (eg, budget risks; risks of project failure; complexity; cybersecurity, etc). The point, however, is that the processes of the NSW AIAF are not *mandated* for any of these systems, so far as we can tell from the brief descriptions in our dataset.

We should also remind the reader that the systems in our dataset are not, in most cases, operating in a fully automated or autonomous fashion. The overwhelming majority of systems identified in the Mapping Project were systems that contributed to the making of decisions that affect NSW residents but involved humans making final decisions.[115] Nevertheless, human involvement at some point is not always the best way to avoid or mitigate potential harms caused by faulty, biased, or otherwise problematic automated systems in high-stakes or sensitive decision-making.[116] This is precisely why the NSW AIAF and the *EU AI Act* (or indeed the Commonwealth government's proposed mandatory guardrails) include human intervention or oversight as only *one* of the governance requirements for AI systems.[117]

V DISCUSSION AND CONCLUSIONS

Policymakers have sought to balance encouraging modernisation and improving services through adoption of advanced digital technologies in the public sector against the well-known, extensively evidenced harms that can arise when public sector decision-making systems are subject to some automation. To achieve this balance, policymakers have developed risk-based regulatory frameworks that impose *ex ante* impact and risk assessment and ongoing monitoring of systems in operation, but only on a subset of systems based on specific criteria. Whilst the precise criteria differ from framework to framework, most involve AI and systems which are considered to pose elevated risks of harms to people and to other important societal interests. This approach has been adopted in two regulatory frameworks examined in this article: the *EU AI Act*, which was passed in 2024, and in the NSW AIAF, originally published in 2022 and revised and updated in July 2024. To define the subset of systems subject to these additional requirements and ensure low-risk systems are not subject to burdensome requirements, policymakers have in both these cases limited additional governance and scrutiny under the regulatory

115 Weatherall et al (n 5).

116 See generally Ben Green, 'The Flaws of Policies Requiring Human Oversight of Government Algorithms' (2022) 45 *Computer Law and Security Review* 105681:1–22 <https://doi.org/10.1016/j.clsr.2022.105681>; Rebecca Crootof, Margot E Kaminski and W Nicholson Price II, 'Humans in the Loop' (2023) 76(2) *Vanderbilt Law Review* 429 <https://dx.doi.org/10.2139/ssrn.4066781>.

117 'NSW AIAF' (n 4) 64; *EU AI Act* (n 3) ch III s 2; 'Safe and Responsible AI in Australia' (n 29) 35–42.

framework to systems that (1) involve the use of AI, and (2) pose an elevated or high risk.

We have sought to evaluate whether these two critical criteria capture what people might consider the 'right' systems: those which pose the clearest risks of harm to people if not developed, deployed, and monitored with care. To examine this question, we have taken advantage of a unique dataset, created as part of a project which aimed to identify as many as possible of the ADM systems in development or used across the entirety of the NSW government. Although created as part of a mapping exercise – and not specifically to enable the kind of analysis undertaken here – this dataset is especially useful precisely because the Mapping Project cast a wide net, deliberately designed to identify systems extending beyond the intended scope of these regulatory frameworks in relation to both conditions. We have, in the dataset:

- Both high- and low-risk systems: the Mapping Project did not impose any risk threshold, other than the basic requirement that the system be involved in decision-making affecting people in NSW (eg, purely 'internal' business systems were excluded); and
- Both systems involving simple automation and traditional computing, *and* systems involving the use of AI.

Looking at 163 systems in use or development across the NSW state government and analysing whether each system (1) would be considered high-risk under the *EU AI Act*, or elevated risk under the NSW AIAF, and (2) would be considered AI according to the EU (and OECD) definition, we found:

(1) While there was considerable overlap in how NSW and EU risk thresholds applied, there were nevertheless a not insignificant number of systems which were considered high/elevated risk under one framework, but not the other;

(2) Perhaps more importantly, there were a significant number of systems that appear to pose elevated risk as set out in the NSW AIAF, but which do *not* involve AI and hence are not subject to the NSW AIAF risk assessment, mitigation and monitoring. Indeed, more systems that seem to pose potentially elevated risk fell outside the NSW AIAF than would appear to be included (29 vs 20 which appear to be covered).

It is important to recognise that our analysis does not reveal anything about how these systems have been designed or developed, or indeed whether they *do* create risks of harm for NSW citizens and residents. But we can say that, by being confined to AI, the NSW AIAF risks failing to ensure the identification and mitigation of risks of harm arising from increased automation and digital technologies for a significant subset of systems being used in the context of sensitive areas and decisions but which do not involve AI.

This analysis assumes a degree of separation between the risk of harm and the technology used. An alternative interpretation of the way frameworks have developed is that the involvement of AI is itself a separate factor that increases the risk of harm to people. That is, it could be argued that the combination of 'risk from AI' and 'risk from other factors' (such as the use case or the kind of decision in which the system is involved) justifies additional scrutiny. This raises

two questions, which again have been the subject of some discussion in the literature and in policy circles: (1) whether 'simple', non-AI automated decision-making, at least in (a subset of) public sector contexts, poses risks that warrant a regulatory framework; and (2) whether there are additional or specific risks arising from the nature of AI that are addressed by a framework like the NSW AIAF; or, alternatively, whether the same or a similar framework should be adopted for both AI and simple forms of ADM, at least in relation to public sector uses.

As to the first of these questions, it is clear that ADM systems used in the public sector have already caused harm to members of the public. This has been most clearly established through the Robodebt Royal Commission, which concluded that an additional regulatory (or rather legislative) framework was required.[118] Within NSW, it has also been established by the NSW Ombudsman's investigation of the Treasury Department's garnishee system, which automatically collected debts to the State from individual bank accounts,[119] and by the Law Enforcement Conduct Commission in relation to the NSW Police Suspect Targeting Management Plan.[120] In short: ADM systems can absolutely cause significant harm to people. The Robodebt Royal Commission found that this justified additional regulatory frameworks; these recommendations were accepted by the government,[121] and a consultation paper has been published by the Commonwealth Attorney-General's Department.[122]

As to the second question, there is significant debate over the question whether AI has characteristics, distinct from other traditional computerised/data-driven systems, warranting a separate or specific regulatory framework. International reports such as the *International AI Safety Report*,[123] and Australian policy documents such as the Commonwealth Government's proposals paper on mandatory guardrails,[124] make the case that AI has specific and additional characteristics warranting special regulatory treatment (such as autonomy, adaptability, and inscrutability), and also amplifies risks already arising from the use of more simple data-driven automated systems (such as the risks arising from relying on biased or low-quality data, speed, and scale). These features may warrant special attention to AI from policymakers.

---

118 *Royal Commission into Robodebt* (n 7) 488 recommendation 17.1. Yee-Fui Ng and Stephen Gray, 'Disadvantage and the Automated Decision' (2022) 43(2) *Adelaide Law Review* 641; Robert van Krieken, 'The Organization of Ignorance: The Australian "Robodebt" Affair, Bureaucracy, Law and Politics' (2024) 50(7–8) *Critical Sociology* 1379 <https://doi.org/10.1177/08969205241245257>; Terry Carney, 'Artificial Intelligence in Welfare: Striking the Vulnerability Balance?' (2020) 46(2) *Monash University Law Review* 23 <https://dx.doi.org/10.2139/ssrn.3805329>.

119 *Revenue NSW Garnishee Order Report* (n 8).

120 Law Enforcement Conduct Commission, *An Investigation into the Use of the NSW Police Force Suspect Targeting Management Plan on Children and Young People: Operation Tepito* (Final Report, October 2023).

121 Australian Government, 'Government Response: Royal Commission into the Robodebt Scheme' (Response, November 2023) 3 ('Government Response to the Royal Commission').

122 'Use of ADM by Government' (n 10).

123 Yoshua Bengio et al, *International AI Safety Report: The International Scientific Report on the Safety of Advanced AI* (Report, January 2025) <https://www.gov.uk/government/publications/international-ai-safety-report-2025> ('*International AI Safety Report*').

124 'Safe and Responsible AI in Australia' (n 29) 11–13.

However, the findings presented here indicate that, for public sector uses, there are strong arguments in favour of extending risk-based frameworks to all ADM. A striking feature of risk-based frameworks like the NSW AIAF is that many if not most of the issues addressed in such frameworks can be applied to simpler automated systems. Simple automated systems may be affected by biases and errors in the data on which they are based, leading to discrimination and unfairness or simply incorrect outcomes. The speed and scale of automated systems can exacerbate the probability and likelihood of harms. Although some of the details (such as of particular technical tests that can be used) might vary, the regulatory frameworks currently being applied to AI systems set out, at a high level, processes such as risk assessment, testing and monitoring, and issues for attention such as data governance and data quality, which if attended to could improve both ADM and AI systems.

There are also practical reasons against limiting regulatory frameworks to AI. It may not always be possible, especially today and increasingly, to determine whether and how AI is being used in data-driven, automated public sector technical systems. Technical systems may include AI without the awareness of users. Simple systems are also perfect candidates for creeping AI in future updates. In fact, the expanded dataset, which includes planned systems, points in that direction. Although the NSW AIAF suggests that such updates should trigger a new self-assessment process, it is legitimate to question whether the need for additional self-assessment will always be apparent to public servants authorising the updates.

Given room for debate about whether a system 'involves AI' or not, it is also important to remember that public servants have incentives to err on the side of non-inclusion.[125] Classifying simpler modes of machine learning as non-AI avoids the requirements of the NSW AIAF (which create workload) and the potential for external scrutiny; although, as we have noted, the NSW AIAF is explicit in encouraging a 'when in doubt, include' approach. We also note that regulating public sector systems only when they involve AI could create perverse incentives *not* to modernise systems, but to stick with simpler, rules-based systems which do not bring in new accountabilities, and, perhaps more problematically, to undermine the adoption of systems that may generate better outputs.[126] We also note that since the Commonwealth government has already accepted the need for frameworks to manage ADM systems post-Robodebt,[127] it would create unnecessary complexities were we to develop separate regulatory frameworks for AI and public sector ADM use.

---

125 See, eg, Jacob Priergaard, 'Not My Debt: The Institutional Origins of Robodebt' (2024) 84(1) *Australian Journal of Public Administration* 142 <https://doi.org/10.1111/1467-8500.12658> (discussing the internal dynamics of bureaucracy); Monika Zalnieriute, 'Against Procedural Fetishism in the Automated State' in Zofia Bednarz and Monika Zalnieriute (eds), *Money, Power and AI: Automated Banks and Automated States* (Cambridge University Press, 2023) 221 <http://dx.doi.org/10.1017/9781009334297> (discussing the 'power' of government).

126 See generally, José-Miguel Bello y Villarino and Ramona Vijeyarasa, 'International Human Rights, Artificial Intelligence, and the Challenge for the Pondering State: Time to Regulate?' (2022) 40(1) *Nordic Journal of Human Rights* 194 <https://doi.org/10.1080/18918131.2022.2069919>.

127 'Government Response to the Royal Commission' (n 121) 21–2.

Applying regulatory frameworks for public sector use to ADM *and* AI focuses attention back on what, arguably, is the more important policy question: the nature and extent of risks of harm, and how to mitigate those risks as we seek to improve, and modernise, the public sector.

It is beyond the scope of this article to develop all of these arguments; many are discussed in the literature. What we hope is that this article adds important new data to the discussion, and a further empirical basis establishing that this question of scope of regulatory frameworks matters. Most importantly, we have put some numbers around the fact that a focus on AI leaves significant numbers of high-risk systems outside of assessment frameworks that could identify and mitigate risks to people in the real world. As such, we hope this is a useful contribution to an active and ongoing policy debate in Australia.