

The Automated Welfare State

Challenges for Socioeconomic Rights of the Marginalised

Terry Carney^{*}

More recently, administrative agencies have introduced ‘new public analytics’ approaches, using data-driven technologies and risk models to reshape how commonplace administrative decisions are produced.¹

5.1 INTRODUCTION

Artificial intelligence (AI) is a broad church. Automated decision-making (ADM), a subset of AI, is the form of technology most commonly encountered in public administration of the social services, a generic term which includes income support (social security) and funding or provision of services such as disability support funding under Australia’s National Disability Insurance Scheme (NDIS). New public analytics is a label that nicely captures how ADM is deployed as the contemporary form of public administration.²

ADM has long been an integral aid to the work of hard-pressed human administrators exercising their delegated social security powers in Centrelink (the specialist service delivery arm of the federal government department called Services Australia). Early digitisation of social security benefits administration not only resulted in considerable efficiency gains but provided the guide-rails that protected against the more egregious errors or decline in decision-making quality as staffing was drastically reduced in scale and shed higher levels skills and experience.

^{*} The author is indebted to research assistance provided by Arundhati Ajith.

¹ Jennifer Raso, ‘Unity in the Eye of the Beholder? Reasons for Decision in Theory and Practice in the Ontario Works Program’ (2020) 70 (Winter) *University of Toronto Law Journal* 1, 2.

² Karen Yeung, ‘Algorithmic Regulation: A Critical Interrogation’ (2018) 12(4) *Regulation & Governance* 505; Lina Dencik and Anne Kaun, ‘Introduction: Datification and the Welfare State’ (2020) 1(1) *Global Perspectives* 12912; Raso, ‘Unity in the Eye of the Beholder?’; Lena Ulbricht and Karen Yeung, ‘Algorithmic Regulation: A Maturing Concept for Investigating Regulation of and through Algorithms’ (2022) 16 *Regulation & Governance* 3.

Automation as such has not been the issue; the issue is a more recent one of a breakneck rush into a ‘digital first future’³ and the abysmal failure of governance, design, ethics, and legal rectitude associated with the \$1.8 billion robodebt catastrophe.⁴ As Murphy J observed in his reasons approving the class action settlement, this was a ‘shameful chapter in the administration of the Commonwealth social security system and a massive failure of public administration [which] should have been obvious to the senior public servants charged with overseeing the Robodebt system and to the responsible Minister at different points’; a verdict echoed by the Royal Commissioner in her July 2023 Report.⁵

ADM is only a technology. Like all new technologies, there are extremes of dystopian and utopian evaluative tropes, though a mature assessment often involves a more ambiguous middle ground.⁶ Like the history of other new technological challenges to law, the answers may call for innovative new approaches, rather than extension of existing remedies. Robodebt was ultimately brought to heel by judicial review and class actions, but the much vaunted ‘new administrative law’ machinery of the 1970s⁷ was seriously exposed. Merits review failed because government ‘gamed it’⁸ while the other accountability mechanisms proved toothless.⁹ So radical new thinking is called for.¹⁰ AI for its part ranges in form from computational aids (or automation) to neural network ‘machine learning’ systems. Even agreed taxonomies of AI are still in development, including recently by the Organisation for Economic Co-operation and Development (OECD), with its four-fold schema of context; data and input; AI model; and task and output.¹¹

The focus of this chapter on social security and social services is apt, because Services Australia (as the former Department of Human Services is now called) was envisaged by the Digital Transformation Agency (‘DTA’ formerly ‘Office’) as ‘the first

³ Terry Carney, ‘Artificial Intelligence in Welfare: Striking the Vulnerability Balance?’ (2020) 46 (2) *Monash University Law Review* 23.

⁴ Tapani Rinta-Kahila et al, ‘Algorithmic Decision-Making and System Destructiveness: A Case of Automatic Debt Recovery’ (2021) 31(3) *European Journal of Information Systems* 313; Peter Whiteford, ‘Debt by Design: The Anatomy of a Social Policy Fiasco – Or Was It Something Worse?’ (2021) 80(2) *Australian Journal of Public Administration* 340.

⁵ *Prygodicz v Commonwealth of Australia (No 2)* [2021] FCA 634, para [5]; Royal Commission into the Robodebt Scheme, Report (Canberra: July 2023).

⁶ Penny Croft and Honni van Rijswijk, *Technology: New Trajectories in Law* (Abingdon, Oxford: Routledge, 2021) 4–16.

⁷ Brian Jinks, ‘The “New Administrative Law”’: Some Assumptions and Questions’ (1982) 41(3) *Australian Journal of Public Administration* 209.

⁸ Joel Townsend, ‘Better Decisions?: Robodebt and Failings of Merits Review’ in Janina Boughey and Katie Miller (eds), *The Automated State* (Sydney: Federation Press, 2021) 52–69.

⁹ Terry Carney, ‘Robo-debt Illegality: The Seven Veils of Failed Guarantees of the Rule of Law?’ (2019) 44(1) *Alternative Law Journal* 4.

¹⁰ Maria O’Sullivan, ‘Automated Decision-Making and Human Rights: The Right to an Effective Remedy’ in Janina Boughey and Katie Miller (eds), *The Automated State* (Sydney: Federation Press, 2021) 70–88.

¹¹ *Framework for the Classification of AI Systems – Public Consultation on Preliminary Findings* (OECD AI Policy Observatory, 2021).

department to roll out intelligent technologies and provide new platforms to citizenry, in accordance with the then DTA's roadmap for later adoption by other agencies'.¹² The focus of this chapter is on the associated risk of digital transformation in the social services, of three main forms. First, the risk due to the heightened vulnerabilities of clients of social services.¹³ Second, the risk from inadequate design, consultation, and monitoring of ADM initiatives in the social services.¹⁴ And finally, the heightened risk associated with particular ADM technologies.

The next section of the chapter (Section 5.2) reviews selected ADM/AI examples in social services in Australia and elsewhere. To draw out differences in levels of risk of various initiatives it takes as a loose organising principle Henman's¹⁵ observation that the risks and pitfalls of AI increase along a progression – lowest where it involves recognising 'patterns', higher where individuals are 'sorted' into categories, and highest where AI is used to make 'predictions'. Section 5.3 discusses the harm inflicted on vulnerable clients of social services when ADM and AI risks are inadequately appreciated, and some options for better regulation and accountability. It questions both the capacity of traditional judicial and administrative machinery in holding AI to account, and the relevance and durability of those 'values' in the face of the transformational *power* of this technology to subordinate and *remake* law and social policy to instead reflect AI values and processes.

Restoration of trust in government is advanced in a short conclusion (Section 5.4) as being foundational to risk management in the social services. Trust is at the heart of the argument made for greater caution, more extensive co-design, and enhanced regulatory oversight of ADM in the social services.

5.2 ISSUES POSED BY AUTOMATION AND ADM

Three issues in particular stand out for social services in Australia. First, the comprehensibility or otherwise of the system for citizens engaging with it. Second, the compatibility or otherwise of ADM in case management. Finally, the risks and benefits of 'predictive' ADM in the social services.

5.2.1 *Comprehensibility Issues*

5.2.1.1 Early Centrelink Adoption of Digitisation and Decision Aids

Prior to robodebt, Centrelink clients concerns mainly centred on intelligibility of digitised social security records and communications, and the ability to understand

¹² Alexandra James and Andrew Whelan, "Ethical" Artificial Intelligence in the Welfare State: Discourse and Discrepancy in Australian Social Services' (2022) 42(1) *Critical Social Policy* 22 at 29.

¹³ Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (New York: St Martins Press, 2017).

¹⁴ Joe Tomlinson, *Justice in the Digital State: Assessing the Next Revolution in Administrative Justice* (Bristol: Policy Press, 2019).

¹⁵ Paul Henman, 'Improving Public Services Using Artificial Intelligence: Possibilities, Pitfalls, Governance' (2020) 42(4) *Asia Pacific Journal of Public Administration* 209, 210.

automation of rate calculations or scoring of eligibility tools. The ADEX and MultiCal systems for debt calculations generate difficult-to-comprehend and acronym-laden print-outs of the arithmetic. This is because the measures were designed for convenience of internal inputting of data rather than ease of consumer comprehension.

The combination of deeply unintelligible consumer documentation and time-poor administrators often leaves too little time to detect less obvious keying or other errors. Internal review officer reconsiderations instead often focus on very basic sources of error such as couple status.¹⁶ While external merits tribunal members do have the skills and expertise to penetrate the fog¹⁷ this only rectifies a very small proportion of such errors (only 0.05% in the case of robodebts), and only for those with the social capital or resources to pursue their concern.

Lack of transparency of communications with run-of-the-mill social security clients remains problematic for want of investment in provision of the ‘public facing’ front-end interfaces (or correspondence templates) to convert an almost 100 per cent digital environment into understandable information for the public. Instead, new investment was initially in pilots to classify and file supporting documents for claims processing.¹⁸ Only in recent years were expressions of interest sought for general customer experience upgrades of the MyGov portal,¹⁹ reinforced by allocation of \$200 million in the 2021–2022 budget for enhancements to provide a ‘simpler and more tailored experience for Australians based on their preferences and interactions’, but also including virtual assistants or chatbots.²⁰

Comprehensibility of debt calculations and other routine high incidence transactions to ordinary citizens surely should be the first reform priority. Transparency to citizens certainly hinges on it. Availability of accurate information to recipients of ADM-based welfare is fundamental to individual due process. This was demonstrated by the contrast between Australia’s failure to explain adequately the basis of yearly income variations under its unlawful ‘robodebt’ calculations, compared to the way case officers in the Swedish student welfare program provided explanations and

¹⁶ Daniel Turner, ‘Voices from the Field’ (Paper presented at the *Automated Decision Making (ADM) in Social Security and Employment Services: Mapping What Is Happening and What We Know in Social Security and Employment Services* (Brisbane, Centre of Excellence for Automated Decision Making and Society (ADM + S), 5 May 2021).

¹⁷ Terry Carney, ‘Automation in Social Security: Implications for Merits Review?’ (2020) 55(3) *Australian Journal of Social Issues* 260.

¹⁸ This is a machine learning optical character reading system developed by Capgemini: Aaron Tan, ‘Services Australia Taps AI in Document Processing’ (16 October 2020) *ComputerWeeklyCom* <www.computerweekly.com/news/252490630/Services-Australia-taps-AI-in-document-processing>.

¹⁹ Sasha Karen, ‘Services Australia Seeks Customer Experience Solutions for myGov Platform Upgrade’ (9 February 2021) ARN <www.arnnet.com.au/article/686126/services-australia-seeks-customer-experience-solutions-mygov-platform-upgrade/>.

²⁰ Asha Barbaschow, ‘All the Tech within the 2021 Australian Budget’ (11 May 2021) *ZDNet* <www.zdnet.com/article/all-the-tech-within-the-2021-australian-budget/>.

an immediate opportunity to rectify inaccurate information.²¹ Even review bodies such as the Administrative Appeals Tribunal (AAT) would benefit from comprehensibility of the basis of decisions. It would benefit from time freed up to concentrate on substantive issues, due to no longer having to pick their way through the morass of computer print-outs and multiple acronyms simply to create an accessible narrative of issues in dispute.²²

5.2.2 ADM Case Management Constraints

5.2.2.1 The (Aborted) NDIS Chatbot

The NDIS is seen as a pathbreaker for digitisation in disability services.²³ But the National Disability Insurance Authority (NDIA) was obliged to abort roll-out of its sophisticated chatbot, called Nadia.

Nadia was designed to assume responsibility for aspects of client interaction and case management. The chatbot was built as a machine learning cognitive computing interface, involving ‘data mining and pattern recognition to interact with humans by means of natural language processing’.²⁴ It was to have an ability to read and adjust to emotions being conveyed, including by lightening the interaction such as by referencing information about a person’s favourite sporting team. However, it did not proceed beyond piloting. As a machine learning system it needed ongoing access to a large training set of actual NDIS clients to develop and refine its accuracy. Rolling it out was correctly assessed as carrying too great ‘a potential risk, as one incorrect decision may disrupt a person’s ability to live a normal life’.²⁵

This risk of error is a serious one, not only for the person affected by it, but also to the confidence of people in public administration. Given the sophistication required of ‘human’ chatbots, it presently must be doubted whether a sufficient standard of performance and avoidance of risk can be attained for vulnerable social security or disability clients. As Park and Humphrey²⁶ suggest, that ability to give human-like cues to end users means that the chatbot ‘need[s] to be versatile and

²¹ Monika Zalnieriute, Lyria Bennett Moses, and George Williams, ‘The Rule of Law and Automation of Government Decision-Making’ (2019) 82(3) *Modern Law Review* 425.

²² Carney, ‘Automation in Social Security’. Marginalised citizens may however benefit from human-centred (a ‘legal design approach’) to AI technologies to broaden access to justice at a relatively low cost: Lisa Toohey et al, ‘Meeting the Access to Civil Justice Challenge: Digital Inclusion, Algorithmic Justice, and Human-Centred Design’ (2019) 19 *Macquarie Law Journal* 133.

²³ Gerard Goggin et al, ‘Disability, Technology Innovation and Social Development in China and Australia’ (2019) 12(1) *Journal of Asian Public Policy* 34.

²⁴ Sora Park and Justine Humphry, ‘Exclusion by Design: Intersections of Social, Digital and Data Exclusion’ (2019) 22(7) *Information, Communication & Society* 934, 944.

²⁵ *Ibid.*, 946.

²⁶ *Ibid.*

adaptable to various conditions, including language, personality, communication style and limits to physical and mental capacities'. This inability of ADM to bridge the 'empathy gap' is why it is so strongly argued that such administrative tasks should remain in the hands of human administrators.²⁷ Even smartphone digital reporting proved highly problematic for vulnerable income security clients such as young single parents under (now abolished) ParentsNext.²⁸ So, it was surely hoping too much to expect better outcomes in the much more challenging NDIS case management environment.

Such issues are not confined to Australia or to case management software of course. Ontario's 'audit trail' welfare management software, deployed to curb a perceived problem of over-generosity, was found to have 'decentred' or displaced caseworkers from their previous role as authoritative legal decision-makers.²⁹ The caseworkers responded by engaging in complicated work-arounds to regain much of their former professional discretion. As Raso concluded, '[s]oftware that requires individuals to fit into pre-set menu options may never be sophisticated enough to deliver complex social benefits to a population as diverse as [Ontario's welfare] recipients'.³⁰

A US federal requirement to automate verification of Medicaid remuneration of disability caregivers provides yet another example. The state of Arkansas adopted an inflexibly designed and user-unfriendly service app (with optional geo-location monitoring). This proved especially problematic for clients who were receiving 'self-directed' care. Care workers were unable to step outside the property boundaries on an errand or to accompany the person without triggering a 'breach' of the service being provided. Unlike Virginia, Arkansas had neglected to take advantage of the ability to exempt self-care, or remove problematic optional elements.³¹

5.2.2.2 NDIA's Aborted ADM Assessment and Planning Reforms

In 2021 public attention was drawn to an NDIA proposal to replace caseworker evaluations by objective rating 'scores' when assessing eligibility for the NDIS, and

²⁷ See Chapter 9 in this book: Cary Coglianese, 'Law and Empathy in the Automated State'.

²⁸ Carney, 'Automation in Social Security'; Simone Casey, 'Towards Digital Dole Parole: A Review of Digital Self-service Initiatives in Australian Employment Services' (2022) 57(1) *Australian Journal of Social Issues* 111. A third of all participants in the program experienced loss or delay of income penalties, with Indigenous and other vulnerable groups overrepresented: Jacqueline Maley, "'Unable to Meet Basic Needs": ParentsNext Program Suspended a Third of Parents' Payments' (11 August 2021) *Sydney Morning Herald* <www.smh.com.au/politics/federal/unable-to-meet-basic-needs-parentsnext-program-suspended-a-third-of-parents-payments-20210811-p58hvl.html>.

²⁹ Jennifer Raso, 'Displacement as Regulation: New Regulatory Technologies and Front-Line Decision-Making in Ontario Works' (2017) 32(1) *Canadian Journal of Law and Society* 75, 83.

³⁰ *Ibid.*, 93.

³¹ Virginia Eubanks and Alexandra Mateescu, "'We Do Not Deserve This": New App Places US Caregivers under Digital Surveillance' (28 July 2021) *Guardian Australia* <www.theguardian.com/us-news/2021/jul/28/digital-surveillance-caregivers-artificial-intelligence>.

to also serve as a basis for providing indicative packages of funding support. This was shelved on 9 July 2021,³² at least in that form.³³ The measure was designed to address inequities around access and size of packages. The stated policy objective was to improve equity of access between different disability groups and between those with and those without access to a good portfolio of recent medical reports, as well as reduce staffing overheads and processing time.³⁴ Subjective assessments of applicant-provided medical reports were to have been replaced by objective ‘scores’ from a suite of functional incapacity ‘tools’. Rating scores were designed not only to improve consistency of NDIS access decisions, but also generate one of 400 personas/presumptive budgets.³⁵

The rating tool and eligibility leg of this reform was not true ADM. That aspect mirrored the historical reform trajectory for Disability Support Pension (DSP) and Carer Allowance/Payments (CA/CP). Originally eligibility for DSP (then called Invalid Pension, IP) was based on showing that an applicant experienced an actual or real life 85 per cent ‘incapacity for work’.³⁶ In the 1990s this was transformed from an enquiry about the real human applicant to becoming an abstraction – assessing the theoretical ability or not of people with that class of functional impairment to be able to perform any job anywhere in the country – and requiring minimum scores under impairment tables rating functional impairment (leaving extremely *narrow* fields/issues for subjective classification of severity). These and associated changes significantly reduced the numbers found eligible for these payments.³⁷ Similar

³² The reforms were opposed by the NDIS Advisory Council and abandoned at a meeting of Federal and State Ministers: Luke Henriques-Gomes, ‘NDIS Independent Assessments Should Not Proceed in Current Form, Coalition’s Own Advisory Council Says’ (8 July 2021) *Guardian Australia* <www.theguardian.com/australia-news/2021/jul/08/ndis-independent-assessments-should-not-proceed-in-current-form-coalitions-own-advisory-council-says>; Muriel Cummins, ‘Fears Changes to NDIS Will Leave Disabled without Necessary Supports’ (7 July 2021) *Sydney Morning Herald* <www.smh.com.au/national/fears-changes-to-ndis-will-leave-disabled-without-necessary-supports-20210706-p58756.html> .

³³ The NDIA outlined significant changes to the model immediately prior to it being halted: Joint Standing C’tte on NDIS, *Independent Assessments* (Joint Standing Committee on the National Disability Insurance Scheme, 2021) 24–27 <https://parlinfo.aph.gov.au/parlInfo/download/committees/reportjnt/024622/toc_pdf/IndependentAssessments.pdf;fileType=application%2Fpdf> .

³⁴ Helen Dickinson et al, ‘Avoiding Simple Solutions to Complex Problems: Independent Assessments Are Not the Way to a Fairer NDIS’ (Melbourne: Children and Young People with Disability Australia, 2021) <<https://apo.org.au/sites/default/files/resource-files/2021-05/apo-nid312281.pdf>> .

³⁵ Ibid; Marie Johnson, ‘“Citizen-Centric” Demolished by NDIS Algorithms’, *InnovationAus* (Blog Post, 24 May 2021) <‘Citizen-centric’ demolished by NDIS algorithms (innovationaus.com)>; Joint Standing C’tte on NDIS, *Independent Assessments* .

³⁶ The original IP test was a subjective one of whether the real applicant with their actual abilities and background could obtain a real job in the locally accessible labour market (if their disability rendered them an ‘odd job lot’ they qualified).

³⁷ Terry Carney, *Social Security Law and Policy* (Sydney: Federation Press, 2006) ch 8; Terry Carney, ‘Vulnerability: False Hope for Vulnerable Social Security Clients?’ (2018) 41(3) *University of New South Wales Law Journal* 783.

changes were made for CA and CP payments. The proposed NDIS assessment tools, distilled from a suite of existing measures and administered by independent assessors (as for DSP), followed the disability payment reform pathway. The risks here were twofold. First that the tool would not adequately reflect the legislative test; second, that the scoring basis would not be transparent or meaningful to clients of the NDIS and their family and advisers.³⁸

The reform did however have a genuine ADM component in its proposed case planning function. The assessment tool was intended not only to determine eligibility for NDIS access but also to then generate one of 400 different ‘template’ indicative funding packages. This leg of the proposed reform was criticised as robo-planning which would result in lower rates of eligibility, smaller and less appropriate packages of support, and loss of individualisation (including loss of personal knowledge reflected in medical reports no longer to be part of the assessment) along with a substantial reduction of human engagement with case planners.³⁹

This was a true deployment of ADM in social services, highlighting Henman’s middle range risks around ADM categorisation of citizens, as well as risks from devaluing professional casework skills, as further elaborated in the next section.

5.2.3 Predictive ADM

Risks associated with ADM are arguably most evident when it is predictive in character.⁴⁰ This is illustrated by the role predictive tools play in determining the level and adequacy of employment services for the unemployed in Australia,⁴¹ and the way compliance with the allocated program of assistance to gain work is tied to retention of eligibility for or the rate of unemployment payments.⁴² The accuracy or otherwise of the prediction is key to both experiences.

5.2.3.1 Predictive ADM Tools in Employment Services and Social Security

Predictive ADM tools to identify those at greatest risk of long-term unemployment operate by allocating people to homogenous bands according to predictors

³⁸ Joint Standing C’tte on NDIS, *Independent Assessments*, ch 5, 9–13.

³⁹ Asher Barbaschow, ‘Human Rights Commission Asks NDIS to Remember Robo-debt in Automation Push’ *ZDNet* (Blog Post, 22 June 2021) <www.zdnet.com/article/human-rights-commission-asks-ndis-to-remember-robo-debt-in-automation-push/>.

⁴⁰ Henman, ‘Improving Public Services Using Artificial Intelligence’, 210.

⁴¹ Mark Considine, Phuc Nguyen, and Siobhan O’Sullivan, ‘New Public Management and the Rule of Economic Incentives: Australian Welfare-to-Work from Job Market Signalling Perspective’ (2018) 20(8) *Public Management Review* 1186.

⁴² Simone Casey, ‘Job Seeker’ Experiences of Punitive Activation in Job Services Australia’ (2022) 57(4) *Australian Journal of Social Issues* 847–60 <<https://doi.org/10.1002/ajs1004.1144>>; Simone Casey and David O’Halloran, ‘It’s Time for a Cross-Disciplinary Conversation about the Effectiveness of Job Seeker Sanctions’ *Austaxpolicy* (Blog Post, 18 March 2021) <www.austaxpolicy.com/its-time-for-a-cross-disciplinary-conversation-about-the-effectiveness-of-job-seeker-sanctions/>.

unemployment duration (statistical profiling). These statistical profiling predictions are much more accurate than random allocation, but still misclassify some individuals. They also fail to identify or account for *causal* reasons for membership of risk bands.⁴³ Human assessments are also liable to misclassify, but professional caseworkers lay claim to richer understandings of causal pathways, which may or may not be borne out in practice.

Predictive tools are constructed in two main ways. As an early pioneer, Australia's Job Seeker Classification Instrument (JSCI) was developed and subsequently adjusted using logistic regression.⁴⁴ Other international designs are constructed using machine learning which interrogates very large data sets to achieve higher accuracy of prediction, as in the Flemish tool.⁴⁵ As with all ADM predictive tools, reflection and reinforcement of bias is an issue: '[b]y definition, high accuracy models trained on historical data to satisfy a bias preserving metric will often replicate the bias present in their training data'.⁴⁶

While there is a large literature on the merits or otherwise of possible solutions for unacceptable bias and discrimination in AI, statistical profiling poses its own quite nuanced ethical challenges. Membership of a racial minority is associated with longer durations of unemployment for instance. But the contribution of racial minority to allocation to a statistical profile band can be either bitter or sweet. Sweet if placement in that band opens a door to *voluntarily* obtaining access to employment services and training designed to counteract that disadvantage (positive discrimination). Bitter if band placement leads to involuntary imposition of requirements to participate in potentially punitive victim blaming programs such as work for the dole. This risk dilemma is real. Thus, a study of the Flemish instrument found that jobseekers not born in the country were 2.6 times more likely to wrongly be classified as at high risk of long-term unemployment.⁴⁷

Nor is the issue confined to the more obvious variables. It arises even with superficially more benign correlations, such as the *disadvantage* actually suffered from having a long duration of employment for a single employer prior to becoming unemployed. Its inclusion in the predictive algorithm is more acceptable if this results in accessing programs to help counter the disadvantage, such as projecting

⁴³ Bert van Landeghem, Sam Desiere, and Ludo Struyven, 'Statistical Profiling of Unemployed Jobseekers' (2021) 483(February) *IZA World of Labor* 5–6 <<https://doi.org/10.15185/izawol.15483>>.

⁴⁴ Sam Desiere, Kristine Langenbucher, and Ludo Struyven, 'Statistical Profiling in Public Employment Services: An International Comparison' (OECD Social, Employment and Migration Working Papers, Paris, OECD Technical Workshop, 2019) 10, 14, 22–23.

⁴⁵ van Landeghem et al, 'Statistical Profiling of Unemployed Jobseekers'.

⁴⁶ Sandra Wachter, Brent Mittelstadt, and Chris Russell, 'Bias Preservation in Machine Learning: The Legality of Fairness Metrics under EU Non-Discrimination Law' (2021) 123(3) *West Virginia Law Review* 735, 775.

⁴⁷ Sam Desiere and Ludo Struyven, 'Using Artificial Intelligence to Classify Jobseekers: The Accuracy-Equity Trade-Off' (2020) 50(2) *Journal of Social Policy* 367.

the human capital benefits of past loyalty to the previous employer compared to likely future sunk costs associated with other applicants with more varied employment histories. But its inclusion is ethically more problematic if it only exposes the person to greater likelihood of incurring income support or other sanctions. Other examples of predictive legal analytics also show that the normative aspect of the law is often supplanted by causal inference drawn from a data set, which may or may not reflect the relevant legal norms.⁴⁸

To a considerable degree, the contribution of statistical profiling hinges on the way it is used. The lack of engagement with causal factors and the arbitrariness or bias of some variables constituting the algorithm is magnified where caseworkers are left with little scope for overriding the initial band allocation. This is the case with Australia's JSCI, a risk compounded by lack of transparency of the algorithm's methodology.⁴⁹ These risks are lessened in employment services systems which leave caseworkers in ultimate control, drawing on assistance from a profiling tool. That is the way the tools are used in Germany, Switzerland, Greece, and Slovenia.⁵⁰

This analysis of the risks associated with predictive tools in employment services is consistent with findings from other areas of law. For example decisions grounded in pre-established facts, such as aspects of aggravating and mitigating criminal sentencing considerations may be more amenable to computation, overcoming perceived deficiencies of instinctive synthesis sentencing law.⁵¹ Distinguishing between administrative decisions as either rule-based or discretionary may prove also useful, because ADM applied to discretionary decisions may result in a failure to lawfully exercise discretion.⁵² Discretionary tasks high in complexity and uncertainty arguably fare better under human supervision and responsibility, such as by a caseworker.⁵³

For its part, Australia mitigates the risk of JSCI predictive errors in two ways. First, an employment services assessment may be conducted by a contracted health or allied health professional in certain circumstances. This is possible where it is shown

⁴⁸ Emre Bayamlioğlu and Ronald Leenes, 'The "Rule of Law" Implications of Data-Driven Decision-Making: A Techno-regulatory Perspective' (2018) 10(2) *Law, Innovation and Technology* 295.

⁴⁹ Jobactive Australia, 'Assessments Guideline – Job Seeker Classification Instrument (JSCI) and Employment Services Assessment (ESAt)' (Canberra: 3 June 2020) <www.dese.gov.au/download/6082/assessments-guideline-job-seeker-classification-instrument-jsci-and-employment-services-assessment/22465/document/pdf>.

⁵⁰ Desiere et al, 'Statistical Profiling in Public Employment Services', 9–10.

⁵¹ Nigel Stobbs, Dan Hunter, and Mirko Bagaric, 'Can Sentencing Be Enhanced by the Use of Artificial Intelligence?' (2017) 41(5) *Criminal Law Journal* 261.

⁵² Justice Melissa Perry, 'AI and Automated Decision-Making: Are You Just Another Number?' (Paper presented at the *Kerr's Vision Splendid for Administrative Law: Still Fit for Purpose? – Online Symposium on the 50th Anniversary of the Kerr Report*, UNSW, 21 October 2021) <www.fedcourt.gov.au/digital-law-library/judges-speeches/justice-perry/perry-j-20211021>.

⁵³ Justin B Bullock, 'Artificial Intelligence, Discretion, and Bureaucracy' (2019) 49(7) *The American Review of Public Administration* 751.

that there are special barriers to employment, a significant change of circumstances or other indications of barriers to employment participation.⁵⁴ The weakness of this is that it occurs only in exceptional circumstances, rather than as part of routine caseworker fine tuning of the overly crude and harsh streaming recommendations resulting from application of the JSCI. So it is essentially confined to operating as a vulnerability *modifier*.

Second, a new payment system has been introduced to break the overly rigid nexus between the JSCI determined stream and the level of remuneration paid to employment providers for a person in that stream. The old rigid payment regime was exposed as perverse both by academic research⁵⁵ and the government's own McPhee Report.⁵⁶ Rather than encourage investment in assisting people with more complex needs it perversely encouraged parking or neglect of such cases in order to concentrate on obtaining greater rewards from assisting those needing little if any help to return to work. The New Enhanced Services model decouples service levels and rates of payment to providers for achieved outcomes, 'which provides some additional flexibility, so a participant with a High JSCI but no non-vocational barriers can be serviced in Tier 1 but still attract the higher outcome payments'.⁵⁷

The obvious question is why Australia's employment services structure and JSCI instrument survived with so little refinement to its fundamentals for nearly two decades after risks were first raised.⁵⁸ Davidson argues convincingly that path-dependence and cheapness were two main reasons why it took until the McPhee Report to effect systemic change.⁵⁹ It is suggested here that another part of the answer lies in a lack of appetite for and difficulty of realising processes of co-design with welfare clients and stakeholders. Certainly, recent experience with co-design in

⁵⁴ DSS, *Guide to Social Security Law* (Version 1.291, 7 February 2022) para 1.1.E.104 <<http://guides.dss.gov.au/guide-social-security-law>> .

⁵⁵ Considine et al, 'New Public Management and the Rule of Economic Incentives'.

⁵⁶ Employment Services Expert Advisory Panel, *I Want to Work* (Canberra: Department of Jobs and Small Business, 2018) <https://docs.jobs.gov.au/system/files/doc/other/final_-_i_want_to_work.pdf>.

⁵⁷ Australia, 'New Employment Services Enhanced Services Payment Model Frequently Asked Questions 13 January 2021' (Web Page, 2023) <www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwjw8K_rxp_vAhVh7nMBHSc8A4YQFjACegQIDBAD&url=https%3A%2F%2Fwww.dese.gov.au%2Fdownload%2F10457%2Fnew-employment-services-enhanced-services-payment-model-frequently-asked-questions%2F19947%2Fnew-employment-services-enhanced-services-payment-model-frequently-asked-questions%2Fpdf&usg=AOvVawoUUO6DiA_ymmo5HEeqjLqR>.

⁵⁸ Mark Considine, *Enterprising States: The Public Management of Welfare-to-Work* (Cambridge: Cambridge University Press, 2001); Terry Carney and Gaby Ramia, *From Rights to Management: Contract, New Public Management and Employment Services* (The Hague: Kluwer Law International, 2002).

⁵⁹ Peter Davidson, 'Is This the End of the Job Network Model? The Evolution and Future of Performance-Based Contracting of Employment Services in Australia' (2022) 57(3) *Australian Journal of Social Issues* 476.

Denmark demonstrates that it is possible to construct a more sophisticated and balanced system which avoids the worst of the adverse effects of statistical profiling and welfare conditionality.⁶⁰ The case for co-production with users is not new to Australian public administration.⁶¹ Co-design is particularly favoured where risks of discrimination and exclusion are present.⁶²

In theory co-design of employment services should also be possible in Australia, but the history of top-down and very harsh 'work-first' welfare-to-work policies⁶³ suggests that its realisation is unlikely.

5.3 RESPONDING TO THE 'POWER' OF AI

[I]n our current digital society, there are three phenomena that simultaneously connect and disconnect citizens from government and impede millions of individuals from exercising their rights on equal terms: bureaucracy, technology, and power asymmetries.⁶⁴

ADM and AI technology in the social services carries a potential both to harm participants as well as to radically transform services by compressing the range of social policy options considered in program design much in the same way these technologies can change the bases of legal accountability (Section 5.3.2).

The power of poorly conceived ADM and AI to inflict unacceptable harm on vulnerable citizens reliant on social services is well established. The challenge here lies in finding ways of mitigating that risk, as discussed below.

5.3.1 *The Vulnerability Challenge in Social Services*

Common to assessing all of these examples of automation and artificial intelligence in welfare is the impact on vulnerable clients. That vulnerability cannot be overstated. As Murphy J wrote in approving the robodebt class action settlement in *Prygodicz*:

⁶⁰ Flemming Larsen and Dorte Caswell, 'Co-Creation in an Era of Welfare Conditionality – Lessons from Denmark' (2022) 51(1) *Journal of Social Policy* 58.

⁶¹ Bill Ryan, 'Co-production: Option or Obligation?' (2012) 71(3) *Australian Journal of Public Administration* 314.

⁶² Joel Tito, BGC Foundation Centre for Public Impact, *Destination Unknown: Exploring the Impact of Artificial Intelligence on Government* (Report, 2017) <www.centreforpublicimpact.org/assets/documents/Destination-Unknown-AI-and-government.pdf>; Elisa Bertolini, 'Is Technology Really Inclusive? Some Suggestions from States Run Algorithmic Programmes' (2020) 20(2) *Global Jurist* 1–76 <<https://doi.org/10.1515/gj-2019-0065>>; Perry, 'AI and Automated Decision-Making'.

⁶³ Simone Casey, 'Social Security Rights and the Targeted Compliance Framework' (2019) February, *Social Security Rights Review* <www.nssm.org.au/social-security-rights-review/social-security-rights-and-the-targeted-compliance-framework/>; Casey, 'Job Seeker Experiences'.

⁶⁴ Sofia Ranchordas and Louisa Scarcella, 'Automated Government for Vulnerable Citizens: Intermediating Rights' (2022) 30(2) *William & Mary Bill of Rights Journal* 373, 375.

It is fundamental that before the state asserts that its citizens have a legal obligation to pay a debt to it, and before it recovers those debts, the debts have a proper basis in law. The group of Australians who, from time to time, find themselves in need of support through the provision of social security benefits is broad and includes many who are marginalised or vulnerable and ill-equipped to properly understand or to challenge the basis of the asserted debts so as to protect their own legal rights. Having regard to that, and the profound asymmetry in resources, capacity and information that existed between them and the Commonwealth, it is self-evident that before the Commonwealth raised, demanded and recovered asserted social security debts, it ought to have ensured that it had a proper legal basis to do so. The proceeding revealed that the Commonwealth completely failed in fulfilling that obligation.⁶⁵

The pain and suffering⁶⁶ from the abysmal failure of governance, ethics, and legal rectitude in the \$1.8 billion robodebt catastrophe⁶⁷ was ultimately brought to heel by judicial review and class actions. Yet as already mentioned, the much vaunted ‘new administrative law’ remedial machinery of the 1970s was seriously exposed. Merits review failed because government ‘gamed it’ by failing to further appeal over 200 adverse rulings that would have made the issue public.⁶⁸ Other accountability mechanisms also proved toothless.⁶⁹ Holding ADM to account through judicial remedies is rarely viable, though very powerful when it is apt.⁷⁰ Judicial review is costly to mount, gameable, and confined to those risks stemming from clear illegality. Robodebt was a superb but very rare exception to the rule, despite the November 2019 settlement victory in the *Amato*⁷¹ test case action, and the sizeable class action compensation settlement subsequently achieved in *Prygodicz*.⁷² A test case launched prior to *Amato* was subject to government litigational gaming. That challenge was halted by the simple step of a very belated exercise of the statutory power to waive the ‘debt’. The same fate could have befallen *Amato* had the then

⁶⁵ *Prygodicz* (No 2), para [7].

⁶⁶ As Murphy J wrote in *Prygodicz* at para [23] ‘One thing, however, that stands out . . . is the financial hardship, anxiety and distress, including suicidal ideation and in some cases suicide, that people or their loved ones say was suffered as a result of the Robodebt system, and that many say they felt shame and hurt at being wrongly branded “welfare cheats”’.

⁶⁷ Whiteford, ‘Debt by Design’.

⁶⁸ Townsend, ‘Better Decisions?’. As pointed out in *Prygodicz*. ‘The financial hardship and distress caused to so many people could have been avoided had the Commonwealth paid heed to the AAT decisions, or if it disagreed with them appealed them to a court so the question as to the legality of raising debts based on income averaging from ATO data could be finally decided’: *Prygodicz* (No 2) para [10].

⁶⁹ Carney, ‘Robo-debt Illegality’.

⁷⁰ Jack Maxwell, ‘Judicial Review and the Digital Welfare State in the UK and Australia’ (2021) 28(2) *Journal of Social Security Law* 94.

⁷¹ *Amato v The Commonwealth of Australia* Federal Court of Australia, General Division, Consent Orders of Justice Davies, 27 November 2019, File No VID611/2019 (Consent Orders).

⁷² *Prygodicz* (No 2).

government been less stubborn in refusing to pay interest on the waived debt.⁷³ For its part, the reasons approving the *Prygodicz* settlement makes clear how remote is the prospect of establishing a government duty of care in negligence, much less establishing proof of breach of any duty of care.⁷⁴

Administrative law judicial or merits review redress predicated on an ‘after-the-event’ interrogation of the *process* of decision-making or the lawfulness (and merits in the case of tribunal review) of the *reasons* for decisions is further undermined by the character of ADM and AI decision-making. This is because neither the decision-making processes followed, nor the thinned down/non-existent reasons generated by the ‘new technological analytics’⁷⁵ are sufficiently amenable to traditional doctrine.⁷⁶ For example, bias arising from the data and code underlying ADM together with biases arising from any human deference to automated outputs, pose evidentiary challenges which may not be capable of being satisfied for the purpose of meeting the requirements of the rule against bias in judicial review.⁷⁷ The ability to bend traditional administrative law principles of due process, accountability, and proportionality to remedy the concerns posed by ADM thus appears to be quite limited.⁷⁸

Outranking all of these concerns, however, is that neither merits review nor judicial review is designed to redress *systemic* concerns as distinct from individual grievances. So radical new thinking is called for,⁷⁹ such as a greater focus on governmentality approaches to accountability.⁸⁰ To understand the gaps in legal and institutional frameworks, the use of ADM systems in administrative settings must be reviewed as a whole – from the procurement of data and design of ADM systems to their deployment.⁸¹ Systemic grievances are not simply a result of purely

⁷³ *Madeleine Masterton v Secretary, Department of Human Services of the Commonwealth* VID73/2019.

⁷⁴ *Prygodicz* (No 2), paras [172]–[183] Murphy J.

⁷⁵ The emerging field of explainable AI (XAI) is a prime example which aims to address comprehension barriers and improve the overall transparency and trust of AI systems. These machine learning applications are designed to generate a qualitative understanding of AI decision-making to justify outputs, particularly in the case of outliers: Amina Adadi and Mohammed Berrada, ‘Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)’ (2018) 6 *IEEE Access* 52138.

⁷⁶ Raso, ‘Unity in the Eye of the Beholder?’.

⁷⁷ Anna Huggins, ‘Decision-Making, Administrative Law and Regulatory Reform’ (2021) 44(3) *University of New South Wales Law Journal* 1048.

⁷⁸ But see: Makoto Cheng Hong and Choon Kuen Hui, ‘Towards a Digital Government: Reflections on Automated Decision-Making and the Principles of Administrative Justice’ (2019) 31 *Singapore Academy of Law Journal* 875; Arjan Widlak, Marlies van Eck, and Rik Peeters, ‘Towards Principles of Good Digital Administration’ in Marc Schuilenburg and Rik Peeters (eds), *The Algorithmic Society* (Abingdon: Routledge, 2020) 67–83.

⁷⁹ O’Sullivan, ‘Automated Decision-Making and Human Rights’, 70–88.

⁸⁰ Raso, ‘Unity in the Eye of the Beholder?’.

⁸¹ Yee-Fui Ng et al, ‘Revitalising Public Law in a Technological Era: Rights, Transparency and Administrative Justice’ (2020) 43(3) *University of New South Wales Law Journal* 1041.

‘mathematical flaws’ in digital systems, as opposed to the product of accountability deficiencies within the bureaucracy and structural injustice.⁸²

One possible new direction is through ADM impact statement processes designed to help prevent systemic grievances. An example is Canada’s Directive, modelled on the GDPR and largely mimicking administrative law values.⁸³ While this certainly has merit, it is open to critique as paying but lip service to risk prevention because it relies on industry collaboration and thus has potential for industry ‘capture’ or other pressures.⁸⁴ Other alternatives include a mixture of ex ante and ex post oversight in the form of an oversight board within the administrative agency to circumvent the barrier of a costly judicial challenge,⁸⁵ and the crafting of sector-specific legal mechanisms.⁸⁶

There is also theoretical appeal in the more radical idea of turning to a governance frame that incorporates administrative accountability norms as its governance standard. The best known of these are Mashaw’s trinity of bureaucratic rationality, moral judgement, and professional treatment, and Adler’s additions of managerialism, consumerist, and market logics.⁸⁷

However these innovative ideas presently lack remedial purchase. Incorporation of tools such as broadened impact assessments may give these norms and values some operational purchase, but the limitations of impact assessment would still remain.⁸⁸ A research impact framework for AI framed around concepts of public value and social value may hold greater promise.⁸⁹

Self-regulation against industry ethics codes, or those co-authored with regulators, has also proven to be weak reeds. They too are easily ‘subsumed by the business logics inherent in the technology companies that seek to self-impose ethical

⁸² Abe Chauhan, ‘Towards the Systemic Review of Automated Decision-Making Systems’ (2020) 25(4) *Judicial Review* 285.

⁸³ Teresa Scassa, ‘Administrative Law and the Governance of Automated Decision-Making: A Critical Look at Canada’s Directive on Automated Decision-Making’ (2021) 54(1) *University of British Columbia Law Review* 251.

⁸⁴ Andrew Selbst, ‘An Institutional View of Algorithmic Impact Assessments’ (2021) 35(1) *Harvard Journal of Law & Technology* 117.

⁸⁵ David Freeman Engstrom and Daniel E Ho, ‘Algorithmic Accountability in the Administrative State’ (2020) 37(3) *Yale Journal on Regulation* 800.

⁸⁶ Frederik J Zuiderveen Borgesius, ‘Strengthening Legal Protection against Discrimination by Algorithms and Artificial Intelligence’ (2020) 24(10) *The International Journal of Human Rights* 1572.

⁸⁷ E.g. Jennifer Raso, ‘Implementing Digitalization in an Administrative Justice Context’ in Joe Tomlinson et al (eds), *Oxford Handbook of Administrative Justice* (Oxford: Oxford University Press, 2021).

⁸⁸ Selbst, ‘An Institutional View of Algorithmic Impact Assessments’.

⁸⁹ Colin van Noordt and Gianluca Misuraca, ‘Evaluating the Impact of Artificial Intelligence Technologies in Public Services: Towards an Assessment Framework’ (Conference Paper, *Proceedings of the 13th International Conference on Theory and Practice of Electronic Governance*, Association for Computing Machinery) 12–15.

codes',⁹⁰ or a form of 'ethics washing'.⁹¹ As Croft and van Rijswijk⁹² detailed for industry behemoths such as Google, this inability to curb corporate power is because it is systemic. As James and Whelan recently concluded:

Codifying ethical approaches might result in better outcomes, but this still ignores the structural contexts in which AI is implemented. AI inevitably operates within powerful institutional systems, being applied to the 'problems' identified by those systems. Digital transformation reinforces and codifies neoliberal agendas, limiting capacities for expression, transparency, negotiation, democratic oversight and contestation . . . This can be demonstrated by juxtaposing the AI ethics discourse in Australia with how AI has been implemented in social welfare.⁹³

The Australian Human Right Commission (AHRC) Report also delivered underwhelming support,⁹⁴ though academic work is continuing to boost the contribution to be made by ethics-based audits.⁹⁵

Consideration of how to mitigate risk of harm to vulnerable recipients of the social services cannot be divorced from meta-level impacts of ADM and AI technology on the character and design of law and social programs, as discussed below.

5.3.2 *The Transformational Power of AI to Shape Social Policy and Law*

Lawyers and social policy designers are rather accustomed to calling the shots in terms of setting normative and procedural standards of accountability (law) and formulating optimally appropriate social service programs (social policy). Digitisation, however, not only transforms the way individual citizens engage with the state and experience state power at the micro-level, but also transforms the nature of government services and modes of governance. The second of these, the transformation of governance by ADM and AI technologies,⁹⁶ is perhaps better known than the first.

Public law scholars have begun to recognise that it may not simply remain a question of how to tame ADM by rendering it accountable to traditional administrative law standards such as those of transparency, fairness, and merits review, but rather of how to avoid those values being supplanted by ADM's values and ways of thinking. The concern is that ADM remakes law in its *technological* image rather than the reverse of making ADM conform to the paradigms of the law.⁹⁷

⁹⁰ Selbst, 'An Institutional View of Algorithmic Impact Assessments', 166.

⁹¹ *Ibid.*, 188.

⁹² Croft and van Rijswijk, *Technology: New Trajectories in Law*, ch 4.

⁹³ James and Whelan, "'Ethical" Artificial Intelligence in the Welfare State', 37.

⁹⁴ Australian Human Rights Commission (AHRC), *Human Rights and Technology: Final Report* (Final Report, 2021) 88–91 <<https://tech.humanrights.gov.au/downloads>>.

⁹⁵ Jakob Mökander et al, 'Ethics-Based Auditing of Automated Decision-Making Systems: Nature, Scope, and Limitations' (2021) 27(4) *Science and Engineering Ethics* 44.

⁹⁶ Fleur Johns, 'Governance by Data' (2021) 17 *Annual Review of Law and Social Science* 4.1.

⁹⁷ Richard Re and Alicia Solow-Niederman, 'Developing Artificially Intelligent Justice' (2019) 22 (Spring) *Stanford Technology Law Review* 242; Carol Harlow and Richard Rawlings,

The same contest between ADM and existing paradigms is evident in other domains of government services. Contemporary advances in design of social services for instance favours ideas such as personalisation, social investment, and holistic rather than fragmented services.⁹⁸ But each of these policy goals is in tension with ADM's design logic of homogenisation and standardisation.⁹⁹ Personalisation of disability services through case planning meetings and devolution of responsibility for individual budgets to clients, in place of top-down imposition of standard packages of services, is one example of that tension, as recently exemplified in the NDIS.¹⁰⁰ The mid-2022 roll-out of algorithmic online self-management of employment services (PEPs) to all except complex or more vulnerable clients is another¹⁰¹ despite introduction of requirement for a digital protection framework under s 159A (7) and (9) of the *Social Security Legislation Amendment (Streamlined Participation Requirements and Other Measures) Act 2022*.

Initiatives across the health and justice systems, such as 'social prescribing' designed to address the contribution of socioeconomic disadvantage to disability and health issues such as by coordinating income support and health

'Proceduralism and Automation: Challenges to the Values of Administrative Law' in Elizabeth Fisher, Jeff King, and Alison Young (eds), *The Foundations and Future of Public Law* (Oxford: Oxford University Press, 2020) 275–98 point out that 'Computerisation is apt to *change the nature of an administrative process*, translating public administration from a person-based service to a dehumanised system where expert systems replace officials and routine cases are handled without human input'.

- ⁹⁸ Australia, *A New System for Better Employment and Social Outcomes* (Final Report, Department of Social Services Reference Group on Welfare Reform to the Minister for Social Services, 2015) <www.dss.gov.au/sites/default/files/documents/02_2015/dss001_14_final_report_access_2.pdf>; Christopher Deeming and Paul Smyth, 'Social Investment after Neoliberalism: Policy Paradigms and Political Platforms' (2015) 44(2) *Journal of Social Policy* 297; Greg Marston, Sally Cowling, and Shelley Bielefeld, 'Tensions and Contradictions in Australian Social Policy Reform: Compulsory Income Management and the National Disability Insurance Scheme' (2016) 51(4) *Australian Journal of Social Issues* 399; Paul Smyth and Christopher Deeming, 'The "Social Investment Perspective" in Social Policy: A Longue Durée Perspective' (2016) 50(6) *Social Policy & Administration* 673.
- ⁹⁹ Jutta Treviranus, *The Three Dimensions of Inclusive Design: A Design Framework for a Digitally Transformed and Complexly Connected Society* (PhD thesis, University College Dublin, 2018) <http://openresearch.ocadu.ca/id/eprint/2745/1/TreviranusThesisVolume1%262_v5_July%204_2018.pdf>; Zoe Staines et al, 'Big Data and Poverty Governance under Australia and Aotearoa/New Zealand's "Social Investment" Policies' (2021) 56(2) *Australian Journal of Social Issues* 157.
- ¹⁰⁰ Terry Carney, 'Equity and Personalisation in the NDIS: ADM Compatible or Not?' a paper delivered at the Australian Social Policy Conference 25–29 October to 1–5 November 2021 Sydney; Alyssa Venning et al, 'Adjudicating Reasonable and Necessary Funded Supports in the National Disability Insurance Scheme: A Critical Review of the Values and Priorities Indicated in the Decisions of the Administrative Appeals Tribunal' (2021) 80(1) *Australian Journal of Public Administration* 97, 98.
- ¹⁰¹ Casey, 'Towards Digital Dole Parole'; Mark Considine et al, 'Can Robots Understand Welfare? Exploring Machine Bureaucracies in Welfare-to-Work' (2022) 51(3) *Journal of Social Policy* 519.

services,¹⁰² or integration of human services and justice systems through justice reinvestment or therapeutic ‘problem-solving’ courts¹⁰³ are two other settings where the same tension arises. In the case of social prescribing, the rigid ‘quantification’ of eligibility criteria for access to the disability pension, together with strict segregation of social security and health services, compounds the issue. In the second instance, predictive criminal justice risk profiling tools threaten to undermine the central rationale of individualisation and flexibility of justice reinvestment interventions to build capacity and avoid further progression into criminality.¹⁰⁴

What is able to be built in social policy terms depends in no small part on the available materials from which it is to be constructed. Rule-based materials such as the algorithms and mechanisms of ADM are unsuited to building social programs reliant on the exercise of subjective discretionary choices. Just as the fiscal objective of reducing staff overheads to a minimum led to enactment of rules in place of former discretionary powers in Australian social security law,¹⁰⁵ government policies such as ‘digital first’ inexorably lead to push back against policies of individualisation and accommodation of complexity. Those program attributes call for expensive professional skills of human caseworkers or the less pricey discretionary judgments of human case administrators. ADM is far less costly than either, so in light of the long reign of neoliberal forms of governance,¹⁰⁶ it is unsurprising that social protection is being built with increasing amounts of ADM and AI,¹⁰⁷ and consequently is sculpted more in the image of that technology than of supposedly favoured welfare policies of personalisation¹⁰⁸ or those of social investment.¹⁰⁹

There are many possible longer-run manifestations should ADM values and interests gain the upper hand over traditional legal values. One risk is that ADM

¹⁰² Alex Collie, Luke Sheehan, and Ashley McAllister, ‘Health Service Use of Australian Unemployment and Disability Benefit Recipients: A National, Cross-Sectional Study’ (2021) 21(1) *BMC Health Services Research* 1.

¹⁰³ Lacey Schaefer and Mary Beriman, ‘Problem-Solving Courts in Australia: A Review of Problems and Solutions’ (2019) 14(3) *Victims & Offenders* 344.

¹⁰⁴ David Brown et al, *Justice Reinvestment: Winding Back Imprisonment* (Basingstoke: Palgrave Macmillan, 2016).

¹⁰⁵ Carney, *Social Security Law and Policy*.

¹⁰⁶ Rob Watts, “Running on Empty”: Australia’s Neoliberal Social Security System, 1988–2015’ in Jenni Mays, Greg Marston, and John Tomlinson (eds), *Basic Income in Australia and New Zealand: Perspectives from the Neoliberal Frontier* (Basingstoke: Palgrave Macmillan, 2016) 69–91.

¹⁰⁷ Monique Mann, ‘Technological Politics of Automated Welfare Surveillance: Social (and Data) Justice through Critical Qualitative Inquiry’ (2020) 1(1) *Global Perspectives* 12991 <<https://doi.org/12910.11525/gp.12020.11299>>.

¹⁰⁸ Andrew Power, Janet Lord, and Allison deFranco, *Active Citizenship and Disability: Implementing the Personalisation of Support*, Cambridge Disability Law and Policy Series (Cambridge: Cambridge University Press, 2013); Gemma Carey et al, ‘The Personalisation Agenda: The Case of the Australian National Disability Insurance Scheme’ (2018) 28(1) *International Review of Sociology* 1.

¹⁰⁹ Smyth and Deeming, ‘The “Social Investment Perspective” in Social Policy’; Staines et al, ‘Big Data and Poverty Governance’.

systems will create subtle behavioural biases in human decision-making,¹¹⁰ changing the structural environment of decision-making. For example the facility of ADM to ascertain and process facts may lead to lesser scrutiny of the veracity of these facts than would be the case in human decision-making. Abdicating the establishment of fact and the value-judgements underlying factfinding to ADM substitutes digital authority for human authority.¹¹¹ This raises questions of accountability where human actors develop automation bias as a result of failing to question outputs generated by an automated system.¹¹²

Other manifestations are more insidious, including entrenchment of an assumption that data-driven decision-making is inherently neutral and objective rather than subjective and contested, or overlooking the contribution of surveillance capitalism discourse around the business practices that procure and commodify citizen data for a profit.¹¹³ This criticism has been levelled at Nordic governmental digitalisation initiatives. The Danish digital welfare state, for example, has drawn academic scrutiny for an apparently immutable belief that data processing initiatives will create a more socially responsible public sector, overlooking the consequences of extensive data profiling using non-traditional sources such as information from individuals' social networking profiles. The public sector's embrace of private sector strategies of controlling consumers through data suggests a propensity for rule of law breaches through data maximisation, invasive surveillance, and eventual citizen disempowerment.¹¹⁴

This is not the place to do other than set down a risk marker about the way ADM and AI may change both the architecture and values of the law as well as of the very policy design of social service programs. That resculpting may be dystopian (less accommodating of human difference and discretions) or utopian in character (less susceptible to chance variability and irrelevant influences known as decisional 'noise'). The reciprocal power contest between the power of AI technology on the one hand and law/social policy on the other is however a real and present concern, as the NDIS example demonstrated.

5.4 TOWARDS AI TRUST AND EMPATHY FOR ORDINARY CITIZENS

Administration of social security payments and the crafting of reasonable and necessary supports under the NDIS are quintessentially examples of how law and government administration impact 'ordinary' citizens. As Raso has observed:

¹¹⁰ Madalina Busuioc, 'Accountable Artificial Intelligence: Holding Algorithms to Account' (2021) 81(5) *Public Administration Review* 825.

¹¹¹ Bertolini, 'Is Technology Really Inclusive?'

¹¹² Busuioc, 'Accountable Artificial Intelligence'.

¹¹³ Shoshana Zuboff, 'Big Other: Surveillance Capitalism and the Prospects of an Information Civilization' (2015) 30(1) *Journal of Information Technology* 75.

¹¹⁴ Rikke Frank Jørgensen, 'Data and Rights in the Digital Welfare State: The Case of Denmark' (2021) 26(1) *Information, Communication & Society* 123–38 <<https://doi.org/10.1080/1369118X.2021.1934069>>.

As public law scholars, we must evaluate how legality or governance functions within administrative institutions in everyday and effectively final decisions. As we develop theories of how it ought to function, we must interrogate how decision making *is* functioning.¹¹⁵

It is suggested here that the principal impression to be drawn from this review of Australia's recent experience of rolling out ADM in Raso's 'everyday' domain of the ordinary citizen, is one of failure of government administration. It is argued that the history so far of Australian automation of welfare – most egregiously the robodebt debacle – demonstrates both a lack of government understanding that the old ways of policy-making are no longer appropriate, and that public trust in government has seriously eroded. Automation of welfare in Australia has not only imposed considerable harm on the vulnerable,¹¹⁶ but has destroyed an essential trust relationship between citizens and government.¹¹⁷

Restoring trust is critical. Trust is one of the five overarching themes identified for consultation in February 2022 by the PM&C's Digital Technology Taskforce and in the AHRC's final report.¹¹⁸ Restoration of trust in the NDIS was also one of the main themes of the recent Joint Parliamentary Committee report on independent assessments.¹¹⁹ Consequently, if future automation is to retain fidelity to values of transparency, quality, and user interests, it is imperative that government engage creatively with the welfare community to develop the required innovative new procedures. A commitment to genuine co-design and collaborative fine-tuning of automation initiatives should be a non-negotiable first step, as stressed for the NDIS.¹²⁰ Ensuring empathy of government/citizen dealings is another.

Emphasising in Chapter 9 about the potential for the automated state, wisely crafted and monitored to realise administrative law values, Cary Coglianese writes that

[i]n an increasingly automated state, administrative law will need to find ways to encourage agencies to ensure that members of the public will continue to have opportunities to engage with humans, express their voices, and receive acknowledgment of their predicaments. The automated state will, in short, also need to be an empathic state.

He warns that '[t]o build public trust in an automated state, government authorities will need to ensure that members of the public still feel a human connection'.

¹¹⁵ Raso, 'Unity in the Eye of the Beholder?.'

¹¹⁶ Carney, 'Artificial Intelligence in Welfare'.

¹¹⁷ Valerie Braithwaite, 'Beyond the Bubble that Is Robodebt: How Governments that Lose Integrity Threaten Democracy' (2020) 55(3) *Australian Journal of Social Issues* 242.

¹¹⁸ AHRC, *Human Rights and Technology*, 24, 28 respectively.

¹¹⁹ Joint Standing C'tte on NDIS, *Independent Assessments*, ix, 22, 120, 152.

¹²⁰ 'Co-design should be a fundamental feature of any major changes to the NDIS': *ibid.*, 145, para 9.28 and recommendation 2.

This calls for a creative new administrative vision able to honour human connection, because '[i]t is that human quality of empathy that should lead the administrative law of procedural due process to move beyond just its current emphasis on reducing errors and lowering costs'. That vision must also be one that overcomes exclusion of the marginalised and vulnerable.¹²¹ Another contribution to building trust is to be more critical of the push for automated administration in the first place. An American 'crisis of legitimacy' in administrative agencies has been attributed to the way uncritical adoption of ADM leads to the loss of the very attributes that justify their existence, such as individualisation.¹²² Framing the NDIS independent assessor episode in this way demonstrated a similar potential deterioration of citizen trust and legitimacy.

Building trust and empathy in social service administration and program design must fully embrace not only the mainstream human condition but also the 'outliers' that AI standardisation excludes.¹²³ At the program design level this at a minimum calls for rejection of any AI or ADM that removes or restricts inclusion of otherwise appropriate elements of personalisation, subjective human judgement, or exercise of discretion relevant to advancing agreed social policy goals. This extends to AI outside the program itself, including being sensitive to indirect exclusion from discriminatory impacts of poorly designed technological tools such as smartphones.¹²⁴

Half a century ago in the pre-ADM 1970s, the 'new administrative law' of merits review and oversight bodies was touted as the way to cultivate citizens' trust in government administration and provide access to administrative justice for the ordinary citizen, though even then the shortfall of preventive avenues was recognised.¹²⁵ Overcoming the ability of government to game first-tier AAT by keeping adverse rulings secret, and arming it with ways of raising systemic issues (such as a form of 'administrative class action') might go a small way to restoring trust and access to justice. But much more creative thinking and work is still to be done at the level of dealing with individual grievances as well.¹²⁶

In short, this chapter suggests that the conversation about the ADM implications for the socioeconomic rights of marginalised citizens in the social services has barely begun. Few remedies and answers currently exist either for program design or for individual welfare administration.

¹²¹ Michael D'Rosario and Carlene D'Rosario, 'Beyond RoboDebt: The Future of Robotic Process Automation' (2020) 11(2) *International Journal of Strategic Decision Sciences (IJSDS)* 1; Jennifer Raso, 'AI and Administrative Law' in Florian Martin-Bariteau and Teresa Scassa (eds), *Artificial Intelligence and the Law in Canada* (Toronto: LexisNexis, 2021).

¹²² Ryan Calo and Danielle Citron, 'The Automated Administrative State: A Crisis of Legitimacy' (2021) 70(4) *Emory Law Journal* 797.

¹²³ Treviranus, *The Three Dimensions of Inclusive Design*.

¹²⁴ Shari Trewin et al, 'Considerations for AI Fairness for People with Disabilities' (2019) 5(3) *AI Matters* 40.

¹²⁵ Jinks, 'The "New Administrative Law"'

¹²⁶ One outstanding question for instance is whether the AHRC Report (AHRC, 'Human Rights and Technology') is correct in thinking that post-ADM merits and judicial review reforms should remain 'technology neutral' or whether more innovative measures are needed.