Computer says "no": automation, algorithms and artificial intelligence in Government decision-making*

Dominique Hogan-Doran SC†

Automating systems can assist individuals making administrative decisions or help identify relevant criteria, evidence or particular issues for consideration. Automation has the potential to make decision-making more accurate, consistent, cost-effective, timely and diminish the risk the decision will be invalidated due to improper motivations or bad faith. However, automated systems raise difficult questions about authorisation and reviewability of decision-making by a non-human agent; there are difficulties applying legislation and resolving complexities and ambiguity; and potential for coding errors. There is still potential for jurisdictional error if there is an error in the automation. This is illustrated by the difficulties associated with the Commonwealth Department of Human Services and Centrelink's online system for raising and recovering social security debts. Machine learning through "Big Data" is a branch of Artificial Intelligence that allows computer systems to learn directly from examples, data and experience. The article analyses the Commonwealth, NSW and SA data-sharing legislation and the need for protection for vulnerable people, such as children or people with a mental illness, transparency and accountability.

Forty years ago, the architects of Australia's administrative review regime contemplated a world in which governmental activity was populated by human policy makers and bureaucrats, not automated decision-making, driven by algorithms, machine learning and artificial intelligence ("AI").[1]

---

* This is a revised and updated version of a paper presented to the UNSW Public Sector Law & Governance Seminar held on 23 May 2017.
The popular phrase "Computer says 'no'" refers to an attitude in customer service in which the default response is to check with information stored or generated electronically and then make decisions based on that, apparently without using common sense, and showing a level of unhelpfulness whereby more could be done to reach a mutually satisfactory outcome, but is not. The phrase gained popularity through the British TV sketch comedy *Little Britain*.

† BEc (SocSc), LLB (Hons), LLM (USyd), BCL (Hons) (Oxon); Senior Counsel, Australian Bar; Adjunct Associate Professor, Faculty of Law, UNSW.

1 AI is often assumed to signify intelligence with fully human capabilities. Such human-level intelligence — or artificial general intelligence — receives significant media prominence, but is still some time from being delivered, and it is unclear when this will be possible. Machine learning is a method that can help achieve "narrow AI", in the sense that many machine learning systems can learn to carry out specific functions "intelligently".

Today, many administrative decisions that used to be based on human reflection are now made automatically. Bureaucracy is no longer what it once was. In the near future, we will live in an "algorithmic society" — one characterised by social and economic decision-making by algorithms and robotic AI agents.[2]

How will public law's mandates of transparency, fairness, and accuracy be guaranteed? When government deprives a citizen of fundamental rights, the principles of natural justice mandate that those citizens are owed notice and a chance to be heard to contest those decisions. Historically, transparency and accountability in administrative decision-making was ensured by the courts, integrity bodies, and the public being able to assess for themselves whether technology promoted lawful decisions. They examined statements of reasons through judicial and merits review, considered information obtained through freedom of information legislation, and followed through on requirements to produce annual reports.

The increasing pervasiveness of automated decision-making makes this public law challenge particularly acute, since the full basis for algorithmic decisions is rarely available to affected individuals:

- the algorithm or some inputs may be secret
- the implementation may be secret, and/or
- the process may not be precisely described.

Only with more governmental transparency, clear and effective regulation, and a widespread awareness of the dangers and mistakes that are already occurring, can we hope to wrest some control of our data from the algorithms that most of us fail to understand.

Automated assistance in administrative decision-making

Automating systems can assist administrative decision-making in a number of ways. For example, they can:

- make the decision
- recommend a decision to the decision-maker
- guide a user through relevant facts, legislation and policy, closing off irrelevant paths as they go

2 In the context of machine learning and AI, a "robot" typically refers to the embodied form of AI; robots are physical agents that act in the real world. These physical manifestations might have sensory inputs and abilities powered by machine learning.

- have capabilities as decision-support systems, providing useful commentary, including about relevant legislation, case law and policy for the decision-maker at relevant points in the decision-making process, and/or
- be used as a self-assessment tool, providing preliminary assessments for individuals or internal decision-makers.

Automated assisted decision-making can help identify:

- the correct question(s) for the decision-maker to determine, including the relevant decision-making criteria and any relevant or irrelevant considerations
- whether any procedures or matters which are necessary preconditions to the exercise of the power have been met or exist
- whether there exists any evidence in respect of each of the matters on which the decision-maker must be satisfied, and
- particular issues which require the decision-maker's consideration and evaluation.

The benefits of automation are not unlimited. Speaking to the Australian Corporate Lawyers Association in 2010,[3] then President of the Administrative Appeals Tribunal, the Honorable Justice Garry Downes, anticipated some of the problems that automated decision-making has generated:

> I have no doubt that in 2020 the unstoppable march of the computer will have continued. Its use in decision-making will have advanced considerably — at all levels, including tribunal and court adjudication. What is important is that this advance should not compromise good administrative decision-making and should not impede or challenge the steady march towards the greatest possible fairness and transparency in that decision-making.
>
> …
>
> This creates three potentially serious, but basic, problems which will affect my canons of good decision-making. First, the wrong data may be entered on the computer. Secondly, the right data may be wrongly entered. In both cases the absence of all the entries on paper makes verification more difficult. Thirdly, the computer may be incorrectly programmed.

3 G Downes, *Looking forward: administrative decision making in 2020*, Australian Corporate Lawyers Association 2010 Government Law Conference, Canberra, 20 August 2010.

Justice Downes was not the first to voice such concerns. In 2004, the Australian Administrative Review Council ("ARC") was one of the first of its kind to consider the growing use of computers to automate governmental decisions. In its Report No 46, *Automated Assistance in Administrative Decision Making*, the ARC examined what kinds of decisions are suitable for automation and how errors can be avoided or, if made, addressed.

The ARC's 27 "best practice" principles for decision-making undertaken with the assistance of automated systems were adopted by the Australian Government Information Management Office's *Automated Decision-Making: Better Practice Guide* in February 2007.[4] The *Better Practice Guide* was directed to automated systems that "build in and automate administrative decision-making *logic* into a computer system":[5]

> Automated systems range from conventional informational technology systems (which may calculate a rate of payment in accordance with a formula set out in legislation) through to more specialised systems such as "expert", "business rules engines", "rules-based" or "intelligent systems" and "decision-support" tools. Business rules engines or rules-based systems (types of expert systems used in administrative decision-making) are software systems that help manage and automate business rules. Generally, these systems contain three main components:
>
> - a knowledge base or rule base containing the relevant business rules (ie legislative, policy or procedural business rules)
> - an independent inference engine which uses reasoning (backward or forward chaining) to draw conclusions, and
> - a user interface which presents questions and information to the user, and supplies the user's response to the inference engine in order to draw conclusions.

A hallmark of this kind of automated system is its ability to examine a set of circumstances (data entered by the user) by applying "business rules" (modelled from legislation, agency policy or procedures) to "decide" dynamically what further information is required, or what choices or information to present to the user, or what conclusion is to be reached.

---

4 *Better Practice Guide*, Appendix B, p 74, available at: www.oaic.gov.au/images/documents/migrated/migrated/betterpracticeguide.pdf, accessed 9 August 2017.

5 ibid p 4.

The *Better Practice Guide* identifies the practice areas that "require particular care with respect to the development and management of automated systems for administrative decision-making",[6] covering the requirements to:

- assess the suitability of automated systems to deliver improved business outcomes for an agency
- establish appropriate project management and governance of automated systems projects
- ensure that the design of an automated system has regard to future requirements (such as maintenance and audit) and complies with privacy legislation
- ensure the continued accuracy of an automated system (including where there are changes to the underlying legislation, policy or procedure)
- ensure the transparency and accountability of the system and its accompanying processes, and
- implement and maintain automated systems appropriately.

A checklist[7] (expressed to be advisory not mandatory) summarised items that should be addressed when considering the implementation or update of an automated system for administrative decision-making.

The *Better Practice Guide* lauds appropriately designed and managed automated systems as offering agencies "potential benefits in consistency, cost efficiency, time economies and new service provision":[8]

> For example:
>
> - they have the potential to make administrative decision-making more accurate and consistent, and it is often in pursuit of these outcomes that agencies consider the automated system option
> - they can offer agencies a cost effective mechanism to make decisions, particularly in policy or program areas where agencies must make many administrative decisions

6 ibid p 7.

7 ibid, summarised in Pt 7, p 57. A more detailed 12 page checklist is available as a "Pocket Guide" at http://finance.gov.au/archive/archive-of-publications/aaadm/docs/AAADM_Pocket_Guide.pdf, accessed 9 August 2017.

8 *Better Practice Guide*, above n 4, p 10.

- they can reduce the time taken for agencies to make an administrative decision,
- new technologies and their application to automated systems can be used in the development of new service delivery options to agencies' customer groups.

In addition to time and cost savings, automated decision-making has the capacity to diminish or even eliminate the risk that a decision will be invalidated by reason of:

- the motivations of the decision-maker (such as decisions made for an improper or ulterior purpose),[9] and/or
- bad faith of the decision-maker (such as decisions made with intended dishonesty, or recklessly or capriciously for an improper or irrelevant purpose, or arbitrarily exceeding power).[10]

The use of automated systems may raise difficult questions about whether Parliament can authorise decision-making by a non-human agent and the reviewability of such decisions. Where a person purports to exercise public decision-making powers, they must be authorised to do so as a matter of law. Equally, if a technological assistant or automated system is being utilised to make part or all of a decision, the use of that system must be authorised. In a paper presented at the Cambridge Centre for Public Law Conference in 2014, Justice Melissa Perry observed that:[11]

> It cannot be assumed that a statutory authority vested in a senior public servant which extends by implication to a properly authorised officer, will also extend to an automated system; nor that authority to delegate to a human decision-maker will permit "delegation" to an automated system. Authority to use such systems should be transparent and express.

---

9 Whether an exercise of power is vitiated by an improper purpose on the part of the decision-maker is determined by reference to his or her subjective state of mind: *Mandurah Enterprises Pty Ltd v WA Planning Commission* (2008) 38 WAR 276 at 289–290; *Austral Monsoon Industries Pty Ltd v Pittwater Council* (2009) 75 NSWLR 169; [2009] NSWCA 154 at [98]. Decisions are impeachable for improper purpose only where the relevant power is purposive or, at least where some purposes are forbidden: M Aronson and M Groves, *Judicial Review of Administrative Action*, 5th ed, Lawbook Co, 2013 at [5.510]. See recently *Golden v V'landys* [2016] NSWCA 300 at [134]–[141].

10 Although arguably not where a decision is vitiated by third party fraud: *SZFDE v Minister for Immigration and Citizenship* (2007) 232 CLR 189; and *SZSXT v Minister for Immigration and Border Protection* (2014) 307 ALR 31; Z Chami, "Fraud in administrative law and the right to a fair hearing" (2010) 61 *Australian Institute of Administrative Law Forum* 5.

11 Subsequently published as M Perry and A Smith, "iDecide: the legal implications of automated decision-making" [2014] *Federal Judicial Scholarship* 17, at www.austlii.edu.au/au/journals/FedJSchol/2014/17.html, accessed 9 August 2017.

Thus, for example, s 495A(1) of the *Migration Act* 1958 (Cth) provides that:

> The Minister may arrange for the use, under the Minister's control, of computer programs for any purposes for which the Minister may, or must, under the designated migration law:
>
> (a) make a decision; or
>
> (b) exercise any power, or comply with any obligation; or
>
> (c) do anything else related to making a decision, exercising a power, or complying with an obligation.

Pursuant to s 495A(2), the Minister is essentially deemed or taken to have made a decision, exercised a power or done something else related to making a decision where that was made, exercised or done by the operation of a computer program. However, as Perry J notes, this raises some unique problems with the concept of "delegating" a decision to an automated system, in whole or in part:[12]

- who is the "decision-maker"?
- to whom has authority been delegated?
  (a) the programmer?
  (b) the policy-maker?
  (c) the authorised decision-maker?
  (d) the computer itself?
- is the concept of delegation appropriately used in this context at all? Unlike human delegates, a computer program can never truly be said to act independently of its programmer or the relevant government agency.
- what if a computer process determines some, but not all, of the elements of the administrative decision? Should the determination of those elements be treated as the subject of separate decisions from those elements determined by the human decision-maker?

The *Better Practice Guide* recognises that governmental agencies may find it difficult to model business rules from legislation and that there may be a need to resolve legislative complexity where ambiguity in interpretation exists.[13] The ARC Report developed this concern, highlighting four areas

12 ibid.

13 *Better Practice Guide*, above n 4, p 36.

in which an error can be made in applying legislation to determine an entitlement:

- the substance and breadth of the legislation (relevant provisions can be found at various locations in a piece of legislation, or across various pieces of legislation)
- the structural complexity of the legislation (eg preconditions can be conjunctive or disjunctive or there may be exceptions to preconditions)
- the semantic complexity of the legislation (certain terms may be difficult to interpret),
- the exercise of discretion (its width, and whether it is being exercised by the decision-maker in accordance with the legislation conferring it).

Automation thus may not eliminate the risk of incorrectly applying statutory requirements, because decisions could be based on a misinterpretation of the applicable legal requirements or an incorrect application of those legal requirements to the facts found by the decision-maker. Laws are interpreted in accordance with statutory presumptions and meaning is also affected by context.[14] One of the greatest challenges is to ensure accuracy in the substantive law applied by automated processes. Through the process of translating laws into code, computer programmers effectively assume responsibility for building decision-making systems. The potential for coding errors is real: as Justice Perry has observed:[15]

> … laws are not static and complex transitional provisions may need to be accommodated, along with relevant common law presumptions. Such systems will need to be kept up to date while maintaining the capacity to apply the law as it stands at previous points in time for decisions caught by transitional arrangements.

In those circumstances, grounds of challenge therefore include that the automated decision-maker:

- identified a wrong issue, asked itself the wrong question, or failed to address the question posed[16]

14 *Project Blue Sky Inc v Australian Broadcasting Authority* (1998) 194 CLR 355 at [69].

15 Perry, above n 11.

16 *Craig v South Australia* (1995) 184 CLR 163; *Minister for Immigration and Multicultural Affairs v Yusuf* (2001) 206 CLR 323 at 351 [82] (McHugh, Gummow, and Hayne JJ).

- ignored relevant material or relied on irrelevant material in a way that affected the exercise of power[17]
- applied a wrong principle of law,[18]
- breached a mandatory statutory procedure or obligation (such as provisions imposing procedural fairness obligations,[19] mandatory time limits, obligations to consult prior to decisions being made, or requiring the giving of reasons for a decision to be valid).[20]

Australian public law has developed a wide range of grounds justifying judicial intervention to overturn administrative decisions or action.[21] The *Better Practice Guide* implicitly recognised there may be a myriad of grounds for challenge, by limiting its guidance on automated systems' treatment of discretion and judgment.[22] The *Better Practice Guide* assumes that discretionary decisions and matters of judgment will be left to human decision-makers, and touches only upon these issues at a high level. In so doing, it acknowledges that the potential benefits of automation are complicated by the substantive expectations of Australian public law:[23]

---

17 *Craig v South Australia*, above. The mere overlooking of, or failing to have regard to, evidence is not a jurisdictional error. Something more is required. In *Minister for Immigration and Citizenship v SZRKT* (2013) 302 ALR 572, Robertson J considered this issue in some detail. His Honour stated at [111]: "the fundamental question must be the importance of the material to the exercise of the tribunal's function and thus the seriousness of any error".

18 That can be shown either from what the decision-maker said, or because the ultimate result, associated with the facts that they expressly or impliedly found, indicates that they must have applied the wrong principle of law: see *Chapman v Taylor* [2004] NSWCA 456 at [33] (Hodgson; Beazley and Tobias JJA agreeing).

19 See eg *Italiano v Carbone* [2005] NSWCA 177 at [106] (Basten JA), [170] (Einstein J) (involved judicial review of a Consumer Trader and Tenancy Tribunal case where damages were made against an entity that was never a party before the tribunal); see also *Duncan v ICAC* [2016] NSWCA 143 at [719] (Basten JA) (one limitation on the powers of ICAC is to be found in the need to accord procedural fairness "before affecting adversely the interests of an individual").

20 *Re Minister for Immigration and Multicultural Affairs; Ex p Lam* (2003) 214 CLR 1.

21 These may be grouped into the following categories: authority to act; application of the law; procedure to be followed; discretion; reasonableness of decision-making; sufficiency of evidence; uncertainty; conduct of the decision-maker (unfair treatment) and motivation of the decision-maker, including unauthorised purpose and bad faith: see C Wheeler, "Judicial review of administrative action: an administrative decision-maker's perspective" (2017) 87 *Australian Institute of Administrative Law Forum* 79 at 81ff.

22 *Better Practice Guide*, above n 4, Pt 5, "Ensuring transparency and accountability", p 43 ff.

23 ibid p 14.

> Despite the potential benefits offered by automated systems, care must be taken to ensure that their use supports administrative law values [of lawfulness, fairness, rationality, openness and efficiency], and that the implementation of an automated system for administrative decision-making will deliver targeted and measurable business benefits to the agency.

This limitation was consistent with the ARC's own caution in setting principles for assessing the suitability of expert systems[24] for administrative decision-making:

- Principle 1: Expert systems that make a decision — as opposed to helping a decision-maker make a decision — would generally be suitable only for decisions involving non-discretionary elements.
- Principle 2: Expert systems should not automate the exercise of discretion.
- Principle 3: Expert systems can be used as an administrative tool to assist an officer in exercising his or her discretion. In these cases the systems should be designed so that they do not fetter the decision-maker in the exercise of his or her power by recommending or guiding the decision-maker to a particular outcome.
- Principle 4: Any information provided by an expert system to assist a decision-maker in exercising discretion must accurately reflect relevant government law and policy.

The *Better Practice Guide* advises that human discretion and judgment be permitted "where relevant", and exhorts system designers to be careful that the system "does not fetter" the decision-maker in exercising any discretion

24 The ARC Report uses the terms "expert systems", whereas the *Better Practice Guide* uses the term "automated systems". See further A Tyree, *Expert systems in law*, Prentice Hall, 1989.

he or she has been given under relevant legislation, policy or procedure.[25] It suggests agencies support human discretion and judgment by:

- outlining and/or breaking down the factors decision-makers should consider when making their judgment,
- providing links to relevant support materials and guides, and
- requiring that decision-makers clearly state and record reasons for decisions, as a statement of reasons or other official (and auditable) output.

Yet even where discretionary decisions are left to humans, deploying automated procedures can still be problematic. Technology-assisted decision-making may frame a decision-maker's consideration of the issues and make determinations of what is relevant and irrelevant. If automated systems are programmed to be used when a discretion or judgment should have been reserved for a human decision-maker, not only would there be a constructive failure to exercise the discretion, but predetermined outcomes may be characterised as pre-judgment or bias.[26]

Generally, only a final, or operative and determinative decision, is administratively reviewable, not the interim steps or conclusions.[27] However, those interim steps may be considered to the extent that they affect the final decision. It follows that, where interim steps in a decision-making process are automated, but the final or operative decision is made by a human, there is the potential for the decision to be affected by a jurisdictional error should there have been an error in the automation process. Similarly, failure by an automated system to bring relevant issues or material to a decision-maker's attention will not absolve the final decision-maker from failing to consider them.[28]

---

25 Fettered decisions includes decisions that were made under the instruction of another person or entity where the decision-maker feels bound to comply; were made when acting on a "purported" delegation which does not permit any discretion as to the decisions to be made; were made under an unauthorised delegation of a discretionary power; involve the inflexible application of a policy without regard to the merits of the particular situation; or improperly fetter the future exercise of statutory discretions, ie a decision-maker with discretionary powers cannot bind itself as to the manner in which those discretionary powers will be exercised in future, whether through a contract, policy, or guideline inflexibly applied: see Wheeler, above n 21, p 82.

26 *Minister for Immigration and Multicultural Affairs v Jia* (2001) 205 CLR 507 at 519 [35], 531–2 [72].

27 *Australian Broadcasting Tribunal v Bond* (1990) 170 CLR 321.

28 *Minister for Aboriginal Affairs v Peko-Wallsend Ltd* (1986) 162 CLR 24.

Critically, automating the process of decision-making will not ipso facto avoid practical injustice, such as may occur if a decision-maker (be they human or automated):

- engages in unfair treatment of persons the subject of the exercise of power[29]
- fails to give notice of the issues in sufficient detail and appropriate time to be able to respond meaningfully
- fails to give a person an opportunity to respond to adverse material that is credible, relevant and significant to the decision to be made
- fails to give access to all information and documents to be relied on,[30]
- misleads a person or entity as to its intention, or fails to adhere to a statement of intention given to a person or entity as to the procedure to be followed, resulting in unfairness.[31]

If an automated assisted decision-maker acts on insufficient evidence, the decision may be open to challenge because it:

- is based on no probative evidence at all[32]

---

29 See *SZRUI v Minister for Immigration, Multicultural Affairs and Citizenship* [2013] FCAFC 80 at [5] (Allsop CJ) ("fair treatment, and apparent fair treatment … involves the recognition of the dignity of the applicant (the subject of the exercise of power) in how the hearing is conducted. That recognition is an inhering element of fairness").

30 For example, *Vega Vega v Hoyle* [2015] QSC 111 (breach of natural justice occurred when the applicant was prevented from having access to information and documents relied on by health service investigators and clinical reviewers in their reports).

31 Because, for example, the person or entity was given no opportunity to be heard in relation to how the process should proceed.

32 For example, the "no evidence rule" of procedural fairness: see eg *Sinclair v Maryborough Mining Warden* (1975) 132 CLR 473; *Australian Broadcasting Tribunal v Bond* (1990) 170 CLR 321 at 355–6 (where a finding made is a critical step in the ultimate conclusion and there is no evidence to support that finding then this may well constitute a jurisdictional error); *SZNKV v Minister for Immigration and Citizenship* (2010) 118 ALD 232 at [38] (Kenny J); *Plaintiff S156/2013 v Minister for Immigration and Border Protection* (2014) 254 CLR 28 at [46] ("[t]he plaintiff also argues that there was no evidence that PNG would fulfil its assurances and would promote the maintenance of a programme which was fair to UMAs [unauthorised maritime arrivals]. However, there was no statutory requirement that the Minister be satisfied of these matters in order to exercise the relevant power. They do not qualify as jurisdictional facts."); *Duncan v ICAC* [2014] NSWSC 1018 at [35(3)] (McDougall J) (declaratory relief may be granted where "there is a finding that is not supported by any evidence whatsoever — that is to say, there is no evidence that could rationally support the impugned finding"); *Duncan v ICAC* [2016] NSWCA 143 at [278], [366] (Bathurst CJ) ("it was an error of law to conclude there was a financial advantage in circumstances where there was no evidence to support the proposition that any financial advantage was obtained by deception"). A no evidence challenge will fail where there is even a slight evidentiary basis to support a finding: see *SZNKV v Minister for Immigration and Citizenship* (2010) 118 ALD 232 at [36]–[38]; *SZUTM v Minister for Immigration and Border Protection* (2016) 149 ALD 317 at [69]–[70].

- is based on a lack of probative evidence, to the extent that there is no basis or is unjustifiable on, or is unsupported by, the available evidence[33] (eg, a "decision which lacks an evident and intelligible justification",[34] decisions "so devoid of any plausible justification that no reasonable body of persons could have reached them",[35] or where there is no evidence to support a finding that is a critical step in reaching the ultimate conclusion)[36]
- is not supported by reasons that "disclose any material by reference to which a rational decision-maker could have evaluated [certain evidence], no such material can be found in the record; and no other logical basis justifies the … finding"[37] (that is, the reasons do not adequately justify the result reached and the court infers from a lack of good reasons that none exist)[38]
- is based on evidence that does not meet the applicable standard of proof;[39] or

33 *Parramatta City Council v Pestell* (1972) 128 CLR 305 at 323 (Menzies J); *Epeabaka v Minister for Immigration and Multicultural Affairs* (1997) 150 ALR 397 (Finkelstein J); *SFGB v Minister for Immigration and Multicultural Affairs* [2003] FCAFC 231 at [19]–[20], [25]; *Australian Pork Ltd v Director of Animal and Plant Quarantine* (2005) 216 ALR 549 at [309]–[310] (Wilcox J). There is no error of law, let alone a jurisdictional error, in the decision-maker making a wrong finding of fact.

34 *Minister for Immigration and Citizenship v Li* (2013) 249 CLR 332 at [76].

35 *Bromley London Borough City Council v Greater London Council* [1983] 1 AC 768 at 821.

36 *SFGB v Minister for Immigration and Multicultural Affairs* [2003] FCAFC 231 at [18]–[20]; *Applicant A227 of 2003 v Minister for Immigration and Multicultural and Indigenous Affairs* [2004] FCA 567 at [44]; *SZJRU v Minister for Immigration and Citizenship* [2009] FCA 315 at [53]–[54].

37 *Minister for Immigration and Citizenship v SZLSP* (2010) 272 ALR 115 at 136 [72].

38 *WAEE v Minister for Immigration and Multicultural and Indigenous Affairs* [2003] FCAFC 184 at [47]; *Minister for Immigration and Border Protection v MZYTS* [2013] FCAFC 114 at [49]–[50]. Note that "inadequate reasons provided at the discretion of the decision-making body cannot impugn the validity of the decision itself": *Obeid v ICAC* [2015] NSWSC 1891 at [49] (Davies J).

39 The appropriate degree of satisfaction may be subject to the "need for caution to be exercised in applying the standard of proof when asked to make findings of a serious nature": see *Sullivan v Civil Aviation Safety Authority* (2014) 322 ALR 581 at [98]–[122] (Flick and Perry JJ). Whether and to what extent this "rule of prudence" should be applied by decision-makers who are not obliged to comply with the "rule" in *Briginshaw v Briginshaw* (1938) 60 CLR 336 because they are not bound by the rules of evidence awaits further examination; see *Sun v Minister for Immigration and Border Protection* [2015] FCCA 2479 at [18]–[20] (Jarrett J); *Bronze Wing International Pty Ltd v SafeWork NSW* [2017] NSWCA 41 at [126]–[127] (Leeming JA).

- is based on insufficient evidence due to inadequate inquiries,[40] including where there is a failure to make reasonable attempts to obtain certain material that is obviously readily available and centrally relevant to the decision to be made.[41]

The evolution of automated decision-making systems: Centrelink and "Robo-Debt"

Recalling Justice Downes's concerns in 2010, errors in computer programming can result in wrong decisions potentially on an enormous scale. For example, if a statutory test is misconstrued when the technology was developed, then that misconstruction could taint any decision made with the assistance of the technology. Automated assisted decision-making may well be productive of deficient reasoning, because the system design leads the (human or automated) decision-maker to:

- give disproportionate/excessive weight to some factor of little importance or any weight to an irrelevant factor or a factor of no importance[42]
- give no consideration to a relevant factor the decision-maker is bound to consider
- fail to base the decision on a rational consideration of the evidence
- use reasoning that is illogical or irrational
- make a decision that lacks an evident and intelligent justification

---

40 *Minister for Immigration and Citizenship v SZIAI* (2009) 83 ALJR 1123; [2009] HCA 39 at [25] ("It may be that a failure to make an obvious inquiry about a critical fact, the existence of which is easily ascertained, could, in some circumstances, supply a sufficient link to the outcome to constitute a failure to review. If so, such a failure could give rise to jurisdictional error by constructive failure to exercise jurisdiction.")

41 The circumstances in which an obligation may be imposed upon an administrator to make further inquiries is repeatedly said to be "strictly limited": see *Prasad v Minister for Immigration and Ethnic Affairs* (1985) 6 FCR 155 at 169–170 (Wilcox J); *Wecker v Secretary, Department of Education Science and Training* (2008) 249 ALR 762 at [109] (per Greenwood J, Weinberg J agreeing). Whether or not it is unreasonable not to make further inquiries may well depend upon the availability of further information and its importance to the factual issues to be resolved. It may also depend upon the subject matter of inquiry and an assessment of the comparative ability of individuals to provide or to obtain relevant information. It is no part of the task of the decision-maker to make out an applicant's case.

42 *Minister for Immigration and Citizenship v Li* (2013) 249 CLR 332 at [72]; *Minister for Aboriginal Affairs v Peko-Wallsend Ltd* (1986) 162 CLR 24 at 30, 41, 71. For additional citations for the propositions advanced in this section of the paper, see Wheeler, above n 21, at pp 82–83.

- base a decision on a mistake in respect of evidence, or on a misunderstanding or misconstruing of a claim advanced by the applicant; or
- make a decision that is contrary to the overwhelming weight of the available evidence.

In this sense, the great efficiency of automated systems could also be their biggest downfall — a lesson illustrated by the Commonwealth Department of Human Services ("DHS") and Centrelink in the launch of a new online compliance intervention ("OCI") system for raising and recovering social security debts in late 2016.

Automated decision-making is authorised by s 6A of the *Social Security (Administration) Act* 1999 (Cth). The OCI system matched the earnings recorded on a customer's Centrelink record with historical employer-reported income data from the Australian Taxation Office ("ATO"). If the customer did not engage with DHS (either online or in person), or if there were gaps in the information provided by the customer, the system filled the gaps with a fortnightly income figure derived from the ATO income data for the relevant employment period ("averaged" data).[43]

The Australian Senate Community Affairs References Committee conducted extensive hearings into the OCI system as part of its Inquiry into the Better Management of the Social Welfare System Initiative.[44] The Committee investigated, inter alia, Centrelink's OCI system and its compliance with debt collection guidelines and Australian privacy and consumer laws, as well as "the impact of Government automated debt collection processes upon the aged, families with young children, students, people with disability and jobseekers and any others affected by the process".

Evidence before the Senate Inquiry suggested that the methodology on which the OCI system was based — that is, utilising a data-matching algorithm that measured income on an annual basis and divided it into

43 See table headed "OCI Information and Statistics" tabled by Department of Human Services, at Canberra public hearing, on 8 March 2017, at www.aph.gov.au/DocumentStore.ashx?id=505d8670-1474-47c5-8d4e-1f72bc3ac183, accessed 11 August 2017.

44 Senate Inquiry into the design, scope, cost-benefit analysis, contracts awarded and implementation associated with the Better Management of the Social Welfare System initiative: at www.aph.gov.au/Parliamentary_Business/Committees/Senate/Community_Affairs/SocialWelfareSystem, accessed 11 August 2017.

equal fortnightly instalments — generated misleading data upon which subsequent decision-making was based.

Witnesses expressed concern that the automatic debt-raising process appeared not to allow for the complexities of casual work, and this was arguably worse if a person is paid considerably in arrears.[45] The system design applied an algorithm that assumed income is distributed evenly over 26 fortnights, with a presumption that work is either full-time or part-time, but with no provision for intermittent, or casual, work.

In its 12 May 2017 submission to the inquiry, Victorian Legal Aid demonstrated how calculating a debt by way of algorithms did not predict an accurate outcome:[46]

> 36. Two recent matters in which VLA has acted have resulted in the AAT referring the debt back to Centrelink to be properly determined, suggesting that calculating a debt by way of algorithms does not provide an accurate outcome.
>
> 37. In the decision of 2016/M103550 (24 March 2017) AAT Member Treble determined that the Tribunal was not satisfied that the debt had been correctly calculated by Centrelink.
>
> 38. At paragraph 17 the Member Treble states that:
>
> > "The relevant income test for Newstart Allowance requires a person's income to be taken into account when it is first earned, derived or received. A fortnightly income test applies. In this case, no effort has been made by Centrelink to obtain actual wage records … even though such records would very likely be readily available if required. Instead it has simply been assumed that the total year earnings can be apportioned equally to each fortnight across the relevant financial year. However, that is not consistent with the requirements of the legislation. The actual pay records are critical to the proper calculation of the overpayment. Accordingly, Centrelink will need to request and obtain those records from the employer in order to arrive at a correct debt calculation."

---

45 See Transcript of Public Hearings at Brisbane, 16 May 2017, pp 2, 28, at www.aph.gov.au/Parliamentary_Business/Committees/Senate/Community_Affairs/SocialWelfareSystem/Public_Hearings, accessed 11 August 2017.

46 Responses to Questions on Notice from the Senate Community Affairs References Committee, Public Hearing, 11 April 2017, at: www.aph.gov.au/DocumentStore.ashx?id=649ce2e1-0521-4c62-bd24-f6a5f3727092, accessed 11 August 2017.

The Commonwealth Ombudsman's Investigation Report published in April 2017 was critical of the debt recovery program generally as to its fairness and reasonableness.[47] Poor service delivery was also a recurring theme in many complaints received by the Ombudsman. Customers had problems getting a clear and transparent explanation about the debt decision and the reasoning behind it.

The design and operation of the Non Employment Income Data Matching ("NEIDM") project, the data-matching process designed to enable DHS to match income data it collects from customers with tax return data reported to the ATO, suffered from a limitation of purpose. The program protocol (which became public in May 2017) made clear that:[48]

> The purpose of the NEIDM program is to identify non-compliant individuals requiring administrative or investigative action. This is identified through the comparison of information held by the Australian Taxation Office (ATO) and what customer[s] have reported to both agencies. The comparison is to identify where DHS customers may have income recorded with the ATO that exceeds the income they have reported to DHS.

Evidence given by Mr Jason McNamara, General Manager, Integrity Modernisation, Department of Human Service, on 16 May 2017[49] gave some insight into the limitations of the data matching program, and improvements made following the Ombudsman's Report:

> Mr McNamara: It is not really just an averaging thing, because data matching is predominantly just saying: "You've got a significant difference. You've told us that you did not earn any income and now we have something from the tax office says you did earn income." That is predominantly what the data matching is showing—it is totals and the differences—
>
> CHAIR: Sorry, but that is not what people have said. We have been shown examples where people have had income and they have declared it.

---

47 Commonwealth Ombudsman, *Centrelink's automated debt raising and recovery system: A report about the Department of Human Services' online compliance intervention system for debt raising and recovery*, April 2017, at www.ombudsman.gov.au/__data/assets/pdf_file/0022/43528/Report-Centrelinks-automated-debt-raising-and-recovery-system-April-2017.pdf, accessed 11 August 2017.

48 A copy of the program protocol is contained in "Answers to written questions on Notice from Senator Kakoschke-Moore, received from Office of the Australian Information Commissioner, 16 May 2017" at www.aph.gov.au/DocumentStore.ashx?id=4f473c96-f345-4314-ad5d-a2c2eecfba47, accessed 11 August 2017.

49 See Transcript of Public Hearings at Brisbane, above n 45, p 46.

> Mr McNamara: The issue in the data matching is better understanding the probability of a debt coming out of that information. That is where we are improving our processes. At the moment we are looking at significant differences, but we also need to take into account, "Are those significant differences likely to lead to a debt or likely to have led to a debt or not?" That is what we are trying to improve at the moment.
>
> There are two sides to that. One is that the information we have is not consistent. But whether the information inconsistency leads to a debt is something we are trying to improve our processes on. That is where we think we can improve. That will help recipients because we will not be sending letters to people who do not have a debt, and it will help us because it will be more efficient for us because we will not be processing recipients who are unlikely to have a debt. From our perspective, if we can increase the probability, that will help everyone. So that is definitely something that is a goal and aim.

The OCI system is a complex automated system that was rolled out on a large scale within a relatively short timeframe. The Ombudsman observed that inevitably there would be problems with the rollout of a system of that scale. The Ombudsman was particularly critical of the planning and risk management work undertaken:[50]

> In our view, many of the OCI's implementation problems could have been mitigated through better project planning and risk management at the outset. This includes more rigorous user testing with customers and service delivery staff, a more incremental rollout, and better communication to staff and stakeholders. […]
>
> A key lesson for agencies and policy makers when proposing to rollout large scale measures which require people to engage in a new way with new digital channels, is for agencies to engage with stakeholders and provide resources for adequate manual support during transition periods. We have recommended DHS undertake a comprehensive evaluation of the OCI in its current form before it is implemented further and any future rollout should be done incrementally.

The Ombudsman expressly cited the 2007 *Better Practice Guide* to reinforce that key considerations in developing automated decision-making systems are whether the system is consistent with administrative law values of lawfulness, fairness, rationality, openness/transparency and efficiency:[51]

---

50 Ombudsman's Report, above n 47, p 3.

51 ibid at p 26 [4.2].

> Customers need to understand how the system works, have the opportunity to present their information in a considered way and be supported in the transition from a manual to an automated system. Clear and comprehensive information to customers and staff is important.

There is always a risk that administrative reviewers will assume a computer system's infallibility. The system audit trail can thus provide critical forensic insights into potential coding and system planning errors that will challenge — or make good — such assumptions.

The data-matching Protocol for the NEIDM project did not identify whether it was developed in accordance with the *Better Practice Guide* or the OAIC *Guidelines on Data Matching in Australian Government Administration*.[52] There was also nothing that identified relevant policy or legislation related to specific income tests that should be applied. Indeed, the NEIDM Protocol only made limited reference to quality controls and audit trails:[53]

> **D. Data quality controls and audit** — When compliance action is proposed, additional checks will take place to ensure the correct DHS customer or spouse has been identified. DHS Customers will be provided with the opportunity to verify the accuracy of the information before any compliance action is taken.
>
> **E. Security and confidentiality** — All DHS computer systems are strictly controlled with features including:
>
> - system access controls and security groupings
> - login identification codes and password protection
> - full audit trails of data files and system accesses.

An audit trail allows for the research and reasoning processes to be comprehended and tracked by a human mind. Automated administrative decisions will be particularly vulnerable to attack where the system does not provide a clear and sufficiently detailed audit trail to make clear what factors have been taken into account. The 2007 *Better Practice Guide* acknowledged this issue in declaring that audit capability is "essential to accountability"[54] and in generating a detailed checklist of measures:

- does the automated system have the capacity to automatically generate a comprehensive audit trail of the administrative decision-making path?
- are all the key decision points identifiable in the audit trail?

---

52 www.oaic.gov.au/agencies-and-organisations/advisory-guidelines/data-matching-guidelines-2014, accessed 11 August 2017.

53 Program protocol, above n 48, p 20.

54 *Better Practice Guide*, above n 4, pp 46–49.

- are all the key decision points within the automated system's logic linked to the relevant legislation, policy or procedure?
- are all decisions recoded and accessible by the system's user, a reviewer or an auditor?
- can the audit trail generated by the automated system be easily integrated into a notification of the decision (including a statement of reasons or other notification) where required?
- is the audit trail secure from tampering (to provide protection and data integrity)?
- does the audit trail include a comprehensive and printable modification history including:
  - who created the document (with time and date recorded)?
  - who has modified the document (with time and date)?
  - a record of what was modified?
  - for privacy and commercial-in-confidence matters, who has viewed the document (with time and date)?
  - who made the final decision (with time and date)?
- does the audit trail start by identifying the authority or delegated authority identified in legislation?
- does the audit trail show who an authorised decision-maker is?
- does the audit trail enable the recording of human intervention in automated processes, for example recording who is authorised to exercise intervention?

The Ombudsman concluded his report by detecting an additional flaw in the system design, which highlighted that a human decision-maker may have exercised differently a discretionary power to waive a 10% recovery fee which was embedded in the business rules encoded in the OCI system:[55]

> 2.36 According to the Administrative Review Council's *Automated Assistance in Administrative Decision Making Better Practice*, a key question in the design of automated decision making systems in administrative law is whether the system is designed "so that the decision-maker is not fettered in the exercise of any discretion or judgement they may have".
>
> 2.37 The *Social Security Act* states that a ten per cent penalty is added to a debt if the debt arose wholly or partly because the person had refused or failed to provide information about their income or had knowingly or recklessly provided incorrect information. However, it also states "this

55 *Ombudsman Report*, ibid n 47, pp 43–44.

section does not apply if the Secretary is satisfied that the person had a reasonable excuse for refusing or failing to provide the information". [Section 1228B *Social Security Act 1991*.]

2.38 The business rules that underpin the application of the reasonable excuse discretion are beneficial if the person engages with the system and indicates there were personal circumstances that impacted their ability to declare their income. This is particularly so in the redesigned system. This means that for people who do engage with the system, the penalty will be manually applied, if at all.

2.39 The penalty will continue to be automatically applied where the department has sought reasonable excuse information, but none has been forthcoming from the customer. If a debt recovery fee is applied, the person will receive a debt notification letter which now provides them with a further opportunity to provide a reasonable excuse and have the fee removed.

2.40 The question of whether these procedural fairness safeguards coupled with the beneficial application of the reasonable excuse provisions are effective in addressing the risk of fettering of the discretion can only be answered by the courts.

2.41 Our observation is that DHS' approach cannot be fair and effective if the department is not effective in its communication to customers about the availability, meaning and importance of reasonable excuse, and the ways of notifying the excuse to the department.

2.42 In the version of the OCI rolled out from 1 July 2016, DHS considered "reasonable excuse" by asking "were there any personal factors that affected your ability to correctly declare your income during the above period/s?". If a person answered "yes" to this question, the penalty fee was not automatically applied by the OCI. If the person answered "no" (or if no answer was provided by the date the debt was raised) the recovery fee was applied automatically.

2.43 In our view, the messaging in the OCI lacked clarity and the "personal circumstances" question may have been insufficient to elicit the necessary reasonable excuse information. In some situations, a person may have answered "no" to the personal circumstances question in situations where a human decision maker, able to review the person's Centrelink record ask relevant questions and consider all the relevant circumstances of the case,

may have decided the penalty fee should not apply, or the discretion not to apply the fee should be exercised. Examples include where:

- income was declared but was not coded into the system because of administrative error
- a customer provided information about fluctuating income on their claim form, but due to administrative error was not placed on fortnightly reporting arrangements
- a customer did not go online or contact DHS (for example, because they thought if the ATO figure was correct they did not need to, or because of vulnerability)
- a customer still believed at the time they answered the question they had declared accurately (note that the question was asked before the customer was notified of the debt) and so did not turn their mind to the question properly
- a customer did not understand what "personal circumstances" meant, or lacked insight into their circumstances
- other situations where information has been provided prior to the intervention.

The final report of the Senate Committee, delivered on 21 June 2017, recorded that in response to the Commonwealth Ombudsman's recommendations, the Department ceased the automatic charging of a 10% debt recovery fee, and now provides information on how individuals can apply not to have this fee imposed where they have a reasonable excuse.[56]

Changing technologies in a changing world: big data, data matching and profiling

In the past, citizens understood that automated systems helped humans apply rules to individual cases. Today, citizens need to appreciate that automated systems can be primary decision-makers, taking human decision-making out of the process. The Senate Committee report concluded that many individuals failed to appreciate the use of automation in the decision-making process, which negatively impacted on their

56 Senate Standing Committee on Community Affairs, *Design, scope, cost-benefit analysis, contracts awarded and implementation associated with the Better Management of the Social Welfare System initiative*, Commonwealth of Australia, 21 June 2017, at [6.17]. A full copy, and chapter links, is available at www.aph.gov.au/Parliamentary_Business/Committees/Senate/Community_Affairs/SocialWelfareSystem/Report, accessed 14 August 2017.

responses to the system's design, misunderstanding the need to engage and supply information when data-matching identified differences:[57]

> When faced with a purported debt, many individuals were unaware of the possibility of an error in the calculations, their right to have a review of that purported debt or how to undertake a review. Many individuals were so daunted by what they saw as an insurmountable task, to challenge a large government department, they simply gave up and paid what they felt was a debt they did not owe.

The technological landscape has substantially changed from that surveyed by the ARC in 2004 and on which the 2007 *Better Practice Guide* was based.[58] They, too, responded to a world in which automated assisted decision-making systems were grounded in logic and rules-based programs that applied rigid criteria to factual scenarios, responding to input information entered by a user in accordance with predetermined outcomes. They focussed on decision *support* tools, not expert systems that could *replace* human discretion and judgment.[59]

In no meaningful way did they contemplate the challenges now posed by machine learning, "Big Data", and the rapidly changing world of data management. Machine learning is a branch of AI that allows computer systems to learn directly from examples, data and experience. Through enabling computers to perform specific tasks intelligently, machine-learning systems can carry out complex processes by learning from data, rather than following pre-programmed rules. As it learns over time, the machine responds to feedback, so that the patterns learnt

---

57 ibid at [6.25].

58 Decision technology, artificial intelligence, data mining and the like were already live issues by 2007. See for example T Davenport and J Harris, "Automated decision making comes of age" *MIT Sloan Management Review*, Summer 2005, 15 July 2005, at: http://sloanreview.mit.edu/article/automated-decision-making-comes-of-age/, accessed 14 August 2017.

59 The *Better Practice Guide* expressly excluded from its definition of automated systems those systems "such as databases (that store information) or case management systems (that track events leading up to an administrative decision being made and/or record decisions once made)". Decision-support applications limited to storing legislative, policy or other information (but which do not automate decision-making logic) were also not considered automated systems for the purposes of the Guide.

yield useful predictions or insights.[60] By increasing our ability to extract insights from ever-increasing volumes of data, machine learning could increase productivity, provide more effective public services, and create new products or services tailored to individual needs.[61]

Machine learning helps extract value from "big data", which is the phenomenon of "high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight, decision making, and process optimization".[62] So called "data mining" (the automatic extraction of implicit and potentially useful information from data) is increasingly used in commercial, scientific and other application areas, powered by increasingly complex Operational Database Management Systems ("ODMS").[63]

Automated decision-making by computer algorithms based on data from human behaviours is becoming fundamental to our digital economy. Such automated decisions can impact everyone, as they now occur routinely not only in government services but:

- health care
- education
- employment
- credit provision.

---

60 Machine learning lives at the intersection of computer science, statistics and data science. An excellent report published on 25 April 2017 by the UK's science academy, the Royal Society, *Machine learning: the power and promise of computers that learn by example,* is available at https://royalsociety.org/~/media/policy/projects/machine-learning/publications/machine-learning-report.pdf, accessed 14 August 2017. Recognising the promise of this technology, in November 2015, the Royal Society launched a policy project on machine learning. This sought to investigate the potential of machine learning over the next 5–10 years, and the barriers to realising that potential. The report's second chapter, on potential applications of machine learning (health care, education, transport and logistics, targeting of public services, finance, pharmaceuticals, energy, retail and the legal sector) is particularly fascinating.

61 ibid p 17.

62 M Beyer and D Laney, *The importance of "big data": a definition*, Gartner, 21 June 2012, at www.gartner.com/id=2057415, accessed 14 August 2017.

63 Encompassing analytical data platforms, scalable Cloud platforms, NoSQL data stores, NewSQL databases, Object databases, Object-relational bindings, graph data stores, service platforms, and new approaches to concurrency control. Roberto Zicari, Professor of Database and Information Systems at Frankfurt University, maintains an educational resource in all these new areas at www.odbms.org/blog, accessed 14 August 2017.

Because public sector data can be a key enabler and catalyse a range of economic activity, Commonwealth and State governments have devised whole of government strategies to deal with big data, its use and sharing. In so doing, governments increasingly appreciate data as an asset, and most recognise it as something to be used for the public good. Thus, the *Australian Public Service Big Data Strategy* released in August 2013[64] expressed that government data is a national asset that should be used for public good:

> Government policy development and service delivery will benefit from the effective and judicious use of big data analytics. Big data analytics can be used to streamline service delivery, create opportunities for innovation, and identify new service and policy approaches as well as supporting the effective delivery of existing programs across a broad range of government operations — from the maintenance of our national infrastructure, through the enhanced delivery of health services, to reduced response times for emergency personnel.

Governments' use of big data and ODMS, like any other form of data or information collection, is subject to a number of legislative controls. Thus, agencies need to comply with the *Data-matching Program (Assistance and Tax) Act* 1990 (Cth) wherever Tax File Numbers are used. The use of big data is also regulated by the *Privacy Act* 1988. At a Commonwealth level, other controls also include:

- *Freedom of Information Act* 1982
- *Archives Act* 1983
- *Telecommunications Act* 1997
- *Electronic Transactions Act* 1999.

In NSW, the *Data Sharing (Government Sector) Act* 2015 removed barriers that impede the sharing of government sector data and implements measures to facilitate the sharing of government sector data with a new Data Analytics Centre and between other agencies. The Data Sharing legislation complements existing NSW legislation. Sharing of personal data is excluded.[65] All data identified in the *Government Information (Public Access)*

---

64 Australian Government Department of Finance and Deregulation, *The Australian Public Service Big Data Strategy*, August 2013, p 5, at www.finance.gov.au/sites/default/files/Big-Data-Strategy.pdf, accessed 14 August 2017.

65 This data is required to be managed according to the *Privacy and Personal Information Protection Act* 1998 and *Health Records and Information Privacy Act* 2002.

Act 2009 as exempt from public release are also specifically exempt from the Data Sharing legislation.[66]

The expansive use of data is acknowledged by the *NSW Government Open Data Policy*,[67] which expresses an aim to release data for use by the community, research sector, business and industry and to accelerate the use of data to derive new insights for better public services. It adopts an open access licensing framework to support the release and reuse of public information, such as by NSW Transport's data exchange program, the Land and Property Information's provision of spatial data, and NSW Education Datahub.

The NSW model contrasts with the legislative environment in South Australia which modelled a *Public Sector (Data Sharing) Act* 2016 on the NSW approach, but without existing privacy legislation or a modern information access regime. The SA legislation takes a different approach by incorporating privacy and a public interest test for sharing in the Trusted Access Principles that govern the provision of information.[68] This approach provides an authorising environment to facilitate both Open Data and data sharing more broadly.

The real value in the use of Big Data lies in its predictive potential. A step beyond data mining, data profiling enables aspects of an individual's personality or behavior, interests and habits to be determined, analysed and predicted.[69] There are vast sources of data used in profiling to build up a picture of an individual. Technologies that generate tracking data

66 These are Sch 1 — Information for which there is a conclusive presumption of overriding public interest against disclosure and Sch 2 — Excluded information of particular agencies.

67 State of New South Wales Department of Finance, Services and Innovation, *NSW Government Open Data Policy*, 2016 at www.finance.nsw.gov.au/ict/sites/default/files/resources/NSW_Government_Open_Data_Policy_2016.pdf, accessed 14 August 2017.

68 *Public Sector (Data Sharing) Act* 2016 (SA), s 7.

69 Article 4(4) of the European General Data Protection Regulation (GDPR) defines "profiling" as "any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements". Available at https://gdpr-info.eu/art-4-gdpr/, accessed 14 August 2017.

(such as smartphones, credit cards, websites, social media and sensors) offer unheralded benefits. Data sources include (but are not limited to):[70]

- education and professional data
- internet search and browsing history
- data derived from existing client/customer relationships
- data collected for credit-worthiness assessments
- financial and payment data
- consumer complaints or queries
- driving and location data
- property ownership data
- information from credit cards and store cards
- consumer buying habits
- wearable technology (such as fitness trackers)
- lifestyle and behavior data gathered from mobile phones
- social network information
- video surveillance systems
- biometric systems
- Internet of Things,
- telematics.

Profiling is not as transparent as other forms of data processing. Profiling can involve predictive elements, which can increase the risk of inaccuracy. Profiling is not always visible, and may take place without an individual's knowledge. Correlations may include hidden biases that have an unintended or discriminatory effect on certain populations. Profiling can emphasise existing stereotypes, social segregation, and limit individual choice and equal opportunities. There is concern that some humans are particularly vulnerable in this area, for example children[71] and those with mental[72] and physical disabilities.

---

70 Information Commissioner's Office, *Feedback request — profiling and automated decision-making*, 6 April 2017, p 3, at https://ico.org.uk/media/about-the-ico/consultations/2013894/ico-feedback-request-profiling-and-automated-decision-making.pdf, accessed 14 August 2017.

71 The European GDPR states that children need particular protection with regard to their personal data. Recital 38 expands "as they may be less aware of the risks, consequences and safeguards concerned and their rights in relation to the processing of personal data. Such specific protection should, in particular, apply to the use of personal data of children for the purposes of marketing or creating personality or user profiles", at https://gdpr-info.eu/recitals/no-38/, accessed 14 August 2017. Controllers must not carry out solely automated processing, including profiling, that produces legal or similar significant effects in respect of a child: GDPR Recital 71, at https://gdpr-info.eu/recitals/no-71/, accessed 14 August 2017.

Profiles also tend to be dynamic and evolving. Profiling may generate new data for an individual based on data relating to other people. Profiling data might also be harvested or mined for information and its commercial value. Profiling where algorithms can discover new correlations may prove useful at a later date.

The following table prepared by the UK Information Commissioner's Office for its April 2017 consultation on *Profiling and Automated-Decision-Making* highlights some of the more widely recognised benefits and risks of profiling:[73]

Benefits	**Risks**
Better market segmentation	Infringement of fundamental rights and freedoms
Permits analysis of risks and fraud	Certain sectors of society may be underrepresented — eg older generation/ vulnerable individuals or those with limited social media presence
Adapting offers of goods and services as well as prices to align with individual consumer demand	Can be used to deduce sensitive personal data from non-sensitive personal data, with a reasonable degree of certainty
Improvements in medicine, education, healthcare and transportation	Unjustifiable deprivation of services or goods
Provide access to credit using different methods to traditional credit-scoring	Risk of data broking industry being set up to use information for their own commercial interest without individual's knowledge
Can provide more consistency in the decision-making process	Using profiling techniques can jeopardise data accuracy

72 See further S Monteith and T Glenn, "Automated decision-making and Big Data: concerns for people with mental illness" (2016) 18(12) *Current Psychiatry Reports* 112, at www.ncbi.nlm.nih.gov/pubmed/27783339, accessed 14 August 2017.

73 Information Commissioner's Office, above n 70, p 6.

The Royal Society in its report published on 27 April 2017, expanded on the social issues associated with machine learning applications:[74]

> As it enhances our analytical capabilities, machine learning challenges our understanding of key concepts such as privacy and consent, shines new light on risks such as statistical stereotyping and raises novel issues around interpretability, verification, and robustness. Some of these arise from the enhanced analytical capabilities provided by machine learning, while others arise from its ability to take actions without recourse to human agency, or from technological issues. While machine learning generates new challenges in these areas, technological advances in machine learning algorithms also offer potential solutions in many cases.
>
> …
>
> Machine learning further destabilises the current distinction between "sensitive" or "personal" and "non-sensitive" data: it allows datasets which at first seem innocuous to be employed in ways that allow the sensitive to be inferred from the mundane.

74 Royal Society, above n 60, p 90.

Internationally, data protection authorities have expressly signaled their intention to closely monitor privacy concerns relating to the use of profiling,[75] Big Data and the evolution of the Internet of Things.[76] In 2018, European member states (including the UK) will introduce new regulations that govern how data usage can be challenged.[77] The *General Data Protection Regulation* ("GDPR") replaces the Data Protection Directive 95/46/EC and is designed to harmonise data privacy laws across Europe. It is focused on the outcome of automated decision-making (which could include profiling) rather than the act of profiling itself.

---

75 *Resolution on Profiling*, 35th International Conference of data protection and privacy commissioners: a compass in turbulent world, Warsaw, Poland, 2013, at https://icdppc.org/wp-content/uploads/2015/02/Profiling-resolution2.pdf, accessed 16 August 2017. The Warsaw Resolution calls upon all parties making use of profiling:

1. To clearly determine the need and the practical use of a specific profiling operation and to ensure appropriate safeguards, before starting with profiling.
2. To limit, consistent with privacy by design principles, the assumptions and the amount of data collected to the level that is necessary for the intended lawful purpose and to ensure that, where appropriate, the data is sufficiently up to date and accurate for its intended purpose.
3. To ensure that the profiles and the underlying algorithms are subject to continuous validation, in order to allow for the improvement of the results and the reduction of false positive or false negative results.
4. To inform society about profiling operations to the maximum extent possible, including the way profiles are assembled and the purposes for which profiles are used, to ensure that individuals are able to maintain control over their own personal data to the maximum extent possible and appropriate.
5. To ensure, in particular with respect to decisions that have significant legal effects on individuals or that affect benefits or status, that individuals are informed about their right to access and correction and that human intervention is provided where appropriate, especially as the predictive power of profiling due to more effective algorithms increases.
6. To ensure that all profiling operations are subject to appropriate oversight.

76 *Resolution Big Data*, 36th International Conference of data protection and privacy commissioners, Mauritius, 2014, at: www.privacyconference2014.org/English/aboutconference/Documents/Resolution/Resolution-Big-Data.pdf, accessed 16 August 2017 and *Mauritius Declaration on the Internet of Things*, 36th International Conference of data protection and privacy commissioners, Mauritius, 2014, at www.privacyconference2014.org/English/aboutconference/Documents/Resolution/Mauritius-Declaration.pdf, accessed 16 August 2017.

77 The European Union ("EU") General Data Protection Regulation ("GDPR") was approved by the EU Parliament on 14 April 2016 and commences operation through a two-year transition period with full implementation and enforcement from 25 May 2018.

Early drafts of the GDPR enshrined what is called a "right to explanation" in law, although research groups argue the final version as approved in 2016 contains no such legal guarantee, as it will depend on interpretation by national and European courts.[78] The GDPR applies to all companies processing and holding the personal data of data subjects residing in the EU, regardless of the company's location. Accordingly, the GDPR will impact globally.

The need for [AI] transparency, [algorithmic] accountability and [robot] ethics

The practical problem with the use of algorithms, especially those used in profiling of human characteristics and behaviours, is that they may be protected by trade secrecy laws, and thus remain impenetrable to outside observers. Algorithms, especially those based on Deep Learning techniques,[79] can also be so opaque that it will be practically impossible to explain how they reach decisions in any event. Indeed, a characteristic of the Artificial Neural Network ("ANN") is that after the ANN has been trained with datasets, any attempt to examine its internal structure to determine why and how it made a particular decision is impossible. The decision-making process of an ANN is and remains opaque. Thus, no-one quite knows how the moves of Google DeepMind's AlphaGo artificial

78 S Wachter, B Mittelstadt, L Floridi, "*Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation*", 2016 International Data Privacy Law (forthcoming), at https://ssrn.com/abstract=2903469, accessed 16 August 2017. This research team hail from the Alan Turing Institute in London and the University of Oxford.

79 See further I Goodfellow, Y Bengio and A Courville, *Deep learning*, MIT Press, 2016 at www.deeplearningbook.org, accessed 18 August 2017. Deep learning (also known as deep structured learning, hierarchical learning or deep machine learning) is a class of machine-learning algorithms that:

- use a cascade of many layers of nonlinear processing units for feature extraction and transformation. Each successive layer uses the output from the previous layer as input. The algorithms may be supervised or unsupervised and applications include pattern analysis (unsupervised) and classification (supervised)
- are based on the (unsupervised) learning of multiple levels of features or representations of the data. Higher level features are derived from lower level features to form a hierarchical representation
- are part of the broader machine learning field of learning representations of data,
- learn multiple levels of representations that correspond to different levels of abstraction; the levels form a hierarchy of concepts.

intelligence program beat the world's top Go player — the feat was "beautiful but puzzling".[80]

This is known as "the black box problem".[81] Within computing, "black box" may describe a situation where we can only observe inputs and outputs; what really happens to the inputs occurs "in the dark". Thus, in completely automated decision-making systems, every input and most outputs are encapsulated in the box too: the information sources representing inputs are not observable. Yet in democracies operating under the rule of law, it is basic that people should enjoy openness and the possibility to make up their own mind and take issue if they disagree with the exercise of governmental powers. Access to information regarding the detailed rules in the system should be publicly available. Can we avoid or manage the "black box problem"?

Because human beings program predictive algorithms, their biases and values are embedded into the software's instructions, known as the source code and predictive algorithms. Algorithms are also what make intelligent robots intelligent. Intelligent robotics blurs the very line between people and instrument.

This new reality is spurring a branch of academic study known as "algorithmic accountability",[82] which focuses on critical questions like:

- how can regulators apply expert judgment given rapidly changing technology and business practices?
- when is human review essential?
- when will controversies over one algorithmic ordering merely result in a second computational analysis of a contested matter?

---

80 J Ribeiro, "AlphaGo's unusual moves prove its AI prowess, experts say", *ComputerWorld*, 14 March 2016, at www.computerworld.com/article/3043457/data-analytics/alphagos-unusual-moves-prove-its-ai-prowess-experts-say.html, accessed 16 August 2017.

81 See F Pasquale, *The black box society: the secret algorithms that control money and information*, Harvard University Press, 2015. Professor Pasquale, a Professor of Law at the University of Maryland, stresses the need for an "intelligible society", one in which we can understand how the inputs that go into these black box algorithms generate the effects of those algorithms. See too N Diakopoulos, *Algorithmic accountability: on the investigation of black boxes*, Tow Center for Digital Journalism, Columbia Journalism School, New York, 2013, at http://towcenter.org/research/algorithmic-accountability-on-the-investigation-of-black-boxes-2/, accessed 30 August 2017.

82 See S Lohr, "If algorithms know all, how much should humans help?", *The New York Times*, 6 April 2015, at https://nyti.ms/1MXHcMW, accessed 23 August 2017. See also D Schartum, "Law and algorithms in the public domain" (2016) *Nordic Journal of Applied Ethics*, 15–26, at http://dx.doi.org/10.5324/eip.v10i1.1973, accessed 16 August 2017.

Some academics argue that "a new concept of technological due process is essential to vindicate the norms underlying last century's procedural protections".[83] Procedural protections could apply not only to the scoring algorithms themselves (a kind of technology-driven rulemaking), but also to individual decisions based on algorithmic predictions (technology-driven adjudication). In their joint work, *The Scored Society: Due Process for Automated Predictions*, law professors Danielle Keats Citron and Frank Pasquale argued:[84]

> Procedural regularity is essential for those stigmatized by "artificially intelligent" scoring systems. The American due process tradition should inform basic safeguards. Regulators should be able to test scoring systems to ensure their fairness and accuracy. Individuals should be granted meaningful opportunities to challenge adverse decisions based on scores miscategorizing them. Without such protections in place, systems could launder biased and arbitrary data into powerfully stigmatizing scores.

Development of artificial intelligence and autonomous systems ("AI/AS") are also giving rise to many complex ethical problems. Although software engineers initially identify the correlations and inferences programmed into algorithms, Big Data promises to eliminate the human middleman at some point in the process. In time, predictive algorithms may evolve to develop an artificial intelligence that guides their own evolution.

---

83 Now Lois K Macht Research Professor of Law, University of Maryland, Danielle Keats Citron first contended in 2008 that a carefully structured inquisitorial model of quality control can partially replace aspects of adversarial justice that automation renders ineffectual: see D Keats Citron, "Technological due process" (2008) 85 *Wash U L Rev* 1249, at http://openscholarship.wustl.edu/law_lawreview/vol85/iss6/2, accessed 16 August 2017; see also D Keats Citron, "Open Code Governance" (2008) 1 *University of Chicago Legal Forum* 9, at http://chicagounbound.uchicago.edu/uclf/vol2008/iss1/9, accessed 16 August 2017.

84 D Keats Citron and F Pasquale, "*The scored society: due process for automated predictions*" (2014) 89:1 *Washington Law Review* 1, at http://digitalcommons.law.umaryland.edu/cgi/viewcontent.cgi?article=2435&context=fac_pubs, accessed 16 August 2017. The use of a "scoring" system in the criminal justice system was highlighted by *Wisconsin v Loomis*, 881 NW 2d 749 (Wis, 2016) where the Supreme Court decided the defendant's right to due process was not violated, despite the trial judge and Circuit Court having referred to the risk assessment score of the defendant in the context of sentencing. The defendant was unable to challenge the process through which the score was reached (it being a trade secret of Northpointe, Inc), although he was given an opportunity to verify some inputs and challenge the overall score, arguing other relevant considerations should be taken into account. See also L Bennett Moses and J Chan, "Using Big Data for legal and law enforcement decisions: testing the new tools" (2014) 37(2) *UNSWLJ* 643.

The most ethically challenging and risky type of decision-making systems are arguably those governmental decisions that:[85]

- have a high impact on peoples' lives;
- maximises the benefit to the decision-maker (to increase revenue, optimise their processes etc) at the expense of the individual; and
- are made on the basis of information over which individuals have no control.

Professor Alan Winfield, a professor of robot ethics at the University of the West of England, accepts that it may be difficult to explain an AI's decision — and for AI, the "black box problem" may well be intractable. Professor Winfield argues (emphasis in original):[86]

> The black box problem may be intractable for ANNs, but could be avoided by using approaches to AI that do not use ANNs.
>
> But — here's the rub. This involves slowing down the juggernaut of autonomous systems and AI development. It means taking a much more cautious and incremental approach, and it almost certainly involves regulation (that, for instance, makes it illegal to run a driverless car unless the car's autopilot has been certified as safe — and *that* would require standards that don't yet exist). Yet the commercial and political pressure is to be **more** permissive, not less; no country wants to be left behind in the race to cash in on these new technologies.
>
> This is why work toward AI/Autonomous Systems standards is so vital, together with the political pressure to ensure our policymakers fully understand the public safety risks of unregulated AI.

The risk that algorithms will make bad decisions that have serious impacts on people's lives has also lead to calls for a supervisory or regulatory body to ensure transparency and fairness. Such a body could have the power to scrutinise and audit algorithms, so they could go in and see whether use

---

85 See further S Finlay, "Ethical risk assessment of automated decision making systems", ODBMS, 2015 at www.odbms.org/2015/02/ethical-risk-assessment-automated-decision-making-systems/, accessed 16 August 2017. Steven Finlay is Head of Analytics at HML, Europe's largest mortgage outsourcing provider. HML is part of the Computershare Group.

86 A Winfield, "The infrastructure of life part 1: safety", *Robohub*, 26 January 2017, at http://robohub.org/the-infrastructure-of-life-part-1-safety/, accessed 30 August 2017. See further, A Winfield and M Jirotka, "The case for an ethical black box", paper presented at the Towards autonomous robotic systems conference, July 2017 at www.researchgate.net/publication/318277040_The_Case_for_an_Ethical_Black_Box, accessed 30 August 2017.

of predictive algorithms which mine personal information to make guesses about individuals' likely actions and risks is actually transparent and fair.[87]

Professor of philosophy Samir Chopra — who discussed the dangers of opaque agents in his 2011 book (with Laurence White) *A legal theory for autonomous artificial agents*[88] — argues that their autonomy from even their own programmers may require them to be regulated as autonomous entities. He suggests an all-encompassing body — a "Federal Robotics Commission" — could deal with the novel experiences and harms robotics

87 I Sample, "AI watchdog needed to regulate automated decision-making, say experts", *The Guardian*, 28 January 2017, at https://www.theguardian.com/technology/2017/jan/27/ai-artificial-intelligence-watchdog-needed-to-prevent-discriminatory-automated-decisions, accessed 16 August 2017.

88 S Chopra and L White, *A legal theory for autonomous artificial agents*, University of Michigan Press, 2011. The authors explain the two views of the goals of artificial intelligence, at p 5:

> From an engineering perspective, as Marvin Minsky noted, it is the "science of making machines do things that would require intelligence if done by men". From a cognitive science perspective, it is to design and build systems that work the way the human mind does. In the former perspective, artificial intelligence is deemed successful along a performative dimension; in the latter, along a theoretical one. The latter embodies Giambattista Vico's perspective of verum et factum convertuntur, "the true and the made are… convertible"; in such a view, artificial intelligence would be reckoned the laboratory that validates our best science of the human mind. This perspective sometimes shades into the claim artificial intelligence's success lies in the replication of human capacities such as emotions, the sensations of taste, and self-consciousness. Here, artificial intelligence is conceived of as building artificial persons, not just designing systems that are "intelligent". [alteration in original.] [citations omitted.]

enables, dedicated to the responsible integration of robotics technologies into society.[89]

Whatever be the type of regulator, the desire to craft a cohesive regulatory theory to underpin algorithmic accountability is finally gathering pace. Professor Frank Pasquale in his recent research paper, *Toward a fourth law of robotics: preserving attribution, responsibility, and explainability in an algorithmic society*[90] builds on the October 2016 lecture by Yale University Law Professor Jack Balkin, *The three laws of robotics in the age of Big Data*, in which Professor Balkin advanced a proposal for a set of "laws of robotics" for an algorithmic society:[91]

1. With respect to clients, customers, and end-users, algorithm users are *information fiduciaries*.
2. With respect to those who are not clients, customers, and end-users, algorithm users have *public duties*. If they are governments, this follows immediately. If they are private actors, their businesses are affected with a public interest. …

89 R Calo, *The case for a federal robotics commission*, Brookings, 2014, at: https://www.brookings.edu/research/the-case-for-a-federal-robotics-commission/, accessed 16 August 2017. See also R Calo, "Robotics and the lessons of cyberlaw" (2015) 103(3) *Calif L Rev* 513 at 513–563 at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2402972, accessed 16 August 2017.

> Robotics combines, for the first time, the promiscuity of data with the capacity to do physical harm; robotic systems accomplish tasks in ways that cannot be anticipated in advance; and robots increasingly blur the line between person and instrument.

Professor Calo is an Assistant Professor at the University of Washington School of Law and former research director at The Center for Internet and Society. He has also been a thought leader in integrating different conceptions of AI to contemporary privacy problems and the field of robotics. See, eg, R Calo, "Robots and privacy", *Robot ethics: the ethical and social implications of robotics*, Patrick Lin et al eds, Cambridge: MIT Press, 2010 at https://ssrn.com/abstract=1599189, accessed 16 August 2017; R Calo, "Open robotics" (2011) 70 *Maryland Law Review* 571; R Calo, "Peeping Hals" 175 (2011) *Artificial Intelligence* 940.

90 F Pasquale, "Toward a fourth law of robotics: preserving attribution, responsibility, and explainability in an algorithmic society", University of Maryland Francis Carey School of Law Legal Studies Research Paper No 2017–21, p 1, at http://ssrn.com/abstract=3002546, accessed 16 August 2017.

91 J Balkin, "The three laws of robotics in the age of Big Data" (2017) 78 *Ohio State Law Journal*, forthcoming, at https://ssrn.com/abstract=2890965, accessed 18 August 2017. Pasquale describes Balkin's lecture as "a tour de force distillation of principles of algorithmic accountability, and a bold vision for entrenching them in regulatory principles": above n 90, p 11.

> 3. The central public duties of algorithm users are not to externalise the costs and harms of their operations. … ["a process of identifying algorithmic nuisance"].[92]

Professor Balkin argues that public law obligations of transparency, due process and accountability flow from these three substantive requirements:[93]

> Transparency — and its cousins, accountability and due process — apply in different ways with respect to all three principles. Transparency and/or accountability may be an obligation of fiduciary relations, they may follow from public duties, and they may be a prophylactic measure designed to prevent unjustified externalization of harms or in order to provide a remedy for harm.

Professor Pasquale identifies the cornerstone of Balkin's proposal as being to create obligations of responsibility in systems that do not necessarily share the human experience of intent. The Royal Society acknowledges that society has "yet to test the boundaries of current models of liability or insurance when it comes to new autonomous intelligent systems",[94] and suggests different approaches to addressing this issue, including:

- the so-called *Bolam* Test,[95] or whether a reasonable human professional would have acted in the same way;
- strict liability — or liability without negligence or intent to harm — for autonomous vehicles; and
- third party liability, akin to provisions made for dangerous dogs.

To deal with the requirement for attribution for the purposes of attributing legal responsibility, Professor Pasquale proposes a fourth law to complement Balkin's first three: "A robot must always indicate the identity of its creator, controller, or owner"[96] (recognising that there may be multiple

---

92 Balkin posits that the best analogy for the harms of algorithmic decision-making is not intentional discrimination but socially unjustified "pollution". Pasquale urges that cost-benefit analysis should "only be one of the methods" used to assess externalities. He also queries whether Balkin's appeal to environmental law principles is sustainable, observing that "we will not always be able to offer precise valuations of the alarm or apprehension we feel at certain algorithmic transformations of human social relations".

93 Balkin, above n 91.

94 Royal Society, above n 60, p 96.

95 *Bolam v Friern Hospital Management Committee* [1957] 1 WLR 582 is an English tort law case that lays down the typical rule for assessing the appropriate standard of reasonable care in negligence cases involving skilled professionals.

96 Pasquale, above n 90, p 10. Pasquale also submits that such a proviso "could also serve as a 'zero-eth' law, complementing the meta-principle that Asimov introduced as his zero-eth law of robotics (namely, that robots must not harm humanity)."

potentially responsible parties for any given machine's development and eventual actions).

Pasquale argues, with some persuasion, that to make Balkin's principles effective, regulators will need to require "responsibility-by-design" to complement extant models of security-by-design and privacy-by-design. That may involve requiring certain hard-coded audit logs in both closed and open robotics, or indeed licences in open robotics that explicitly contemplate problematic outcomes. Initiatives like these will not simply regulate robotics post hoc, but necessarily influence systems development, by foreclosing some design options, and encouraging others.

One computer science researcher has already generated a general framework for "accountable" algorithms, unifying a suite of tools from cryptography to design processes that enable after-the-fact oversight, consistent with the norm in law and policy.[97] The work demonstrates that it is possible for accountable algorithms to attest to the valid operation of a decision policy even when all or part of that policy is kept secret.

Professor Winfield, mentioned above, is also leading development of a new standard on transparency in autonomous systems, based on the simple principle that it should always be possible to find out why an AI or robot made a particular decision. Professor Winfield heads a British Standards Institute working group on robot ethics to develop industry standards for AIs that aims to make them transparent and more accountable.[98] The group drafted a new standard BS 8611 *Guide to the ethical design of robots and robotic systems*, published in April 2016, believed to be the world's first standard on ethical robots. Also in 2016, the very well regarded IEEE standards association — the same organisation that gave us WiFi — launched a *Global initiative on ethical considerations in AI and autonomous systems.*[99] In December 2106, the IEEE published *Ethically aligned design: a vision for prioritising human wellbeing with AI and autonomous systems.*[100] The purpose of this initiative is to ensure every technologist is educated and empowered to prioritise ethical considerations in the design and

97 J Kroll, "Accountable algorithms", Academic dissertation (PhD) Princeton University, 2015, at http://arks.princeton.edu/ark:/88435/dsp014b29b837r, accessed 16 August 2017.

98 Professor Winfield maintains a very interesting blog "Mostly, but not exclusively, about robots" at http://alanwinfield.blogspot.com.au, accessed 16 August 2017.

99 See http://standards.ieee.org/develop/indconn/ec/ead_law.pdf, accessed 16 August 2017.

100 See http://standards.ieee.org/develop/indconn/ec/ead_v1.pdf, accessed 16 August 2017.

development of autonomous and intelligent systems; in a nutshell, to ensure ethics is "baked in".

Ethical issues almost always directly translate into concrete legal challenges — or they give rise to difficult collateral legal problems. A challenge for policy-makers will be to develop an approach to assessing the ethical risks associated with automated decision-making systems that will be simple and pragmatic, easily incorporated as part of a standard risk assessment exercise, and can be undertaken during a system's design phase.

Appropriate action can then help to mitigate the latent risk, for example, by:

- undertaking analysis to identify "at risk" groups that may not be treated fairly (ethically) by the system (eg ethnic minorities with poor literacy or language skills, children and people suffering from mental illness or physical disability as might impede their interaction);
- designing constraints and over-ride rules to ensure that "at risk" groups are treated in a fair way; and
- after the system goes live, monitoring the situation on a regular basis, so that constraints and over-rides can be fine-tuned as required.

Conclusion

Intelligent robotics is shaping up to be the next transformative technology of our times. Their increasing role in our lives merits systematic reassessment and changes to law, institutions, and the legal profession and academy.[101] Critical evaluation is necessary to avoid addressing policy questions in a piecemeal fashion, with increasingly poor outcomes, and slow accrual of knowledge. The literature on the difficult theoretical questions of AI ethics, algorithmic system ethics, and big data and ethics, is now vast, well-established and easily accessible. The legal profession, judiciary and academy should be well placed to support an informed public debate about what we want algorithms, machine learning, and AI to do, and how the benefits can best be distributed.

101 See further R Susskind and D Susskind, *The future of the professions: how technology will transform the work of human experts*, Oxford University Press, 2015.