International Journal of Law and Information Technology, 2020, 28, 193–216
doi: 10.1093/ijlit/eaaa016
Article

A statutory right to explanation for decisions generated using artificial intelligence

Joshua Gacutan* and Niloufer Selvadurai †

ABSTRACT

As artificial intelligence technologies are increasingly deployed by government and commercial entitles to generate automated and semi-automated decisions, the right to an explanation for such decisions has become a critical legal issue. As the internal logic of machine learning algorithms is typically opaque, the absence of a right to explanation can weaken an individual's ability to challenge such decisions. This article considers the merits of enacting a statutory right to explanation for automated decisions. To this end, this article begins by considering a theoretical justification for a right to explanation, examines consequentialist and deontological approaches to protection and considers the appropriate ambit of such a right, comparing absolute transparency with partial transparency and counterfactual explanations. This article then analyses insights provided by the European Union's General Data Protection Regulation before concluding by recommending an option for reform to protect the legitimate interests of individuals affected by automated decisions.

KEYWORDS artificial intelligence, algorithmic transparency, automated decision-making, privacy, right to explanation, General Data Protection Regulation

INTRODUCTION

As government and commercial entities increasingly use artificial intelligence (AI) technologies to automate decisions, a critical issue to be addressed is whether there should be a right to explanation for such decisions. 'Automated decisions' are decisions based on AI which are generated by machine learning algorithms that analyse large amounts of personal and historic data to make evaluative assessments about individuals.[1] The degree of automation varies, ranging from fully automated decisions to semi-automated decisions that involve differing degrees and forms of human intervention and oversight. Entities around the world are relying on automated decisions due to their ability to make decision-making more efficient and

* Joshua Gacutan, Ashurst Solicitors, E-mail: joshua.gacutan1@gmail.com
† Niloufer Selvadurai, Macquarie Law School Macquarie University Sydney Australia, E-mail: niloufer.selvadurai@mq.edu.au

1 Bryan Casey, Ashkan Farhangi and Roland Vogl, 'Rethinking Explainable Machines: The GDPR's "Right to an Explanation" Debate and the Rise of Algorithmic Audits in Enterprise' (2019) 34 Berkeley Technology Law Journal 145, 147.

consistent, at a scale and cost incomparable to that of human decision-making.[2] Decisions that were historically made by humans, such as eligibility for a welfare payment, eligibility for a bank loan or the assessment of an income tax credit, are now increasingly automated.[3] However, as the internal logic of automated decision-making systems is often opaque[4] and unavailable to the public,[5] affected individuals are commonly not provided with an explanation of the underlying rationale for the decision.[6]

In such a context, the right to explanation envisaged in the European Union's 2018 General Data Protection Regulation (GDPR)[7] forms an important development in this field. While the legal status of this right to explanation has been the subject of considerable debate, the GDPR allows an individual to seek 'meaningful information' about the 'logic' involved in making a decision, in circumstances where the decision was made solely using automated technologies and the decision produced legal effects concerning the individual or significantly affected them.[8] To this end, the objective of this article is to analyse the merits of enacting a statutory right to explanation for automated decisions in Australia.

This article begins by identifying the regulatory challenges presented by automated decision-making and considers how it can undermine an individual's right to privacy, control and dignity. Building on this analysis, this article puts forward a theoretical justification for the enactment of a right to explanation, examining the differences between consequentialist and deontological approaches to protection, and considers the appropriate ambit of such a right, comparing absolute transparency with partial transparency and analysing the value of counterfactual explanations. This article then progresses to analysing the applicability of existing Australian laws and considers how the present limitations in such laws could be addressed by enacting a right to explanation similar to that enacted by the European Union GDPR. While the legal status of this GDPR right to explanation has been the subject of debate, the framing of the right provides valuable insights into potential reforms in comparative jurisdictions. The article concludes by proposing a right to explanation for automated decision-making in Australia that seeks to calibrate the societal efficiency gains of automated decision-making with the protection of legitimate personal rights and interests.

2 ibid 149.

3 See, for example, Danielle Citron and Frank Pasquale, 'The Scored Society: Due Process for Automated Predictions' (2014) 89 Washington Law Review 1, 4; Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard UP 2015) 3.

4 See, for example, Sandra Wachter, Brent Mittelstadt and Chris Russell, 'Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR' (2018) 31 Harvard Journal of Law & Technology 841, 843.

5 Jenna Burrell, 'How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms' (2016) 3 Big Data and Society 1, 10.

6 Wachter, Mittelstadt and Russell (n 4) 843.

7 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L119, 1.

8 art 22(1) of the GDPR; See Casey, Farhangi and Vogl (n 1) 147.

REGULATORY CHALLENGES PRESENTED BY AUTOMATED DECISION-MAKING

The development and nature of automated decision-making systems

Automated decision-making systems have developed in line with the first two waves of AI, computer systems that function as if possessing human intelligence.[9] The father of modern computing, Alan Turing, first suggested that AI could be identified if a human interrogating the responses from both a computer and another human were unable to determine which response was produced by the computer.[10] The first wave of AI refers to rule-based computer systems developed in the 1970s.[11] These systems involve explicit human-authored rules which are built into a computer system and then applied to new scenarios to reach the pre-empted conclusion.[12] In the context of decision-making, first wave AI systems are able to automate decisions based on clear, fixed and finite criteria.[13] If legislation, for example, stipulates that individuals who meet certain criteria are eligible for a welfare payment, a first wave automated decision-making system would operate so that only individuals meeting the specified criteria would receive the payment.[14]

The second wave of AI refers to computer systems that create their own rules.[15] Unlike the explicit human-authored rules of first wave AI systems, second wave AI systems use machine learning algorithms that construct their own rules through analysing patterns and correlations in personal and historic data.[16] In this way, second wave AI systems are capable of continually 'learning' new information and identifying more complex patterns in data.[17] In the USA, the Conference of Chief Justices expressed their support for the use of the second wave automated decision-making system called Correctional Offender Management Program for Alternative Sanctions (COMPAS).[18] COMPAS employs machine learning algorithms to analyse defendants' personal data, historic public data, criminal history and interviews to determine which defendants pose the highest risk of re-offending.[19] In *Wisconsin v Loomis*,[20] for example, the Wisconsin Supreme Court ordered a standard pre-sentencing investigation report which involved the use of COMPAS to analyse data from surveys, interviews and public records to determine whether Mr Loomis was likely to re-offend.[21]

9 Alan Turing, 'Computing Machinery and Intelligence' (1950) 59 Mind 433, 440.
10 ibid.
11 See, for example, Alan Tyree, *Expert Systems in Law* (Prentice Hall 1989).
12 See Monika Zalnieriute, Bennett Moses and George Williams, 'The Rule of Law and Automation of Government Decision-Making' (2019) University of New South Wales Law Report Series 14.
13 ibid.
14 ibid.
15 ibid.
16 ibid.
17 Zalnieriute, Moses and Williams (n 12) 436; See also David Lehr and Paul Ohm, 'Playing with the Data: What Legal Scholars Should Learn about Machine Learning' (2017) 51 University of California Davis Law Review 653, 669.
18 CCJ/COSCA Criminal Justice Committee, In Support of the Guiding Principles on Using Risk and Needs Assessment Information in the Sentencing Process (Resolution 7, 2011).
19 Zalnieriute, Moses and Williams (n 12) 437.
20 (2016) 881 NW 2d 749 (Wisconsin).
21 ibid.

The decision-making process can be automated to varying extents depending on whether a first wave or second wave automated decision-making system is employed.[22] However, both forms of decision-making systems do not remove humans from the process of automating decisions.[23] Computer programmers still determine what data science techniques to employ, what categories of data are collected and which data variables will form the criteria for the automated decisions.[24] Most of the automation comes after computer programmers have designed and built the automated decision-making system.[25]

Concerns as to the lack of transparency and loss of individual control

Automated decision-making systems lack transparency and present a threat to an individual's dignity and control as they make evaluations about individuals without revealing the rationale for such decisions.[26] Ananny and Crawford argue that transparency is not simply a 'precise end state in which everything is clear' to individuals, but rather a system of observing and understanding that promises a form of control.[27] In this vein, individuals feel more secure when automated decisions are transparent as the notion of seeing something will necessarily lead to control over it.[28] If automated decisions, for example, are transparent individuals, it will have greater control over the processing of their personal data which may lead to a more accurate collection of their personal data.[29] Furthermore, Schwartz argues that the dignity of individuals requires individuals to understand and know the rationale behind automated decisions.[30] If individuals are denied an understanding of how decisions about them are made, they will lose self-worth, and over time the legitimacy of the automated decision-making system will be questioned because of the lack of understanding and loss of dignity and control.[31]

A variety of recent American cases have examined the legality of decisions generated using AI. In *Houston Federation of Teachers et al. v Houston ISD*,[32] the Houston Independent School District used AI software to assess the performance and impact of their teachers by analysing the test results of their respective classes.[33] The results of the AI software were then used to substantiate the dismissal of teachers deemed

22 Raja Parasuraman and Victor Riley, 'Humans and Automation: Use, Misuse, Disuse, Abuse' (1997) 39 Human Factors 230, 232.

23 Zalnieriute, Moses and Williams (n 12) 435.

24 ibid.

25 ibid.

26 Mike Ananny and Kate Crawford, 'Seeing Without Knowing: Limitations of the Transparency Ideal and its Application to Algorithmic Accountability' (2016) 20 New Media & Society 973, 975; Paul Schwartz, 'Data Processing and Government Administration: The Failure of the American Legal response to the Computer' (1992) 43 Hastings Law Journal 1321, 1348.

27 Mike Ananny and Kate Crawford, 'Seeing Without Knowing: Limitations of the Transparency Ideal and its Application to Algorithmic Accountability' (2016) 20 New Media & Society 973, 975.

28 ibid.

29 Danielle Citron, 'Technological Due Process' (2008) 85 Washington University Law Review 1249, 1253.

30 Paul Schwartz, 'Data Processing and Government Administration: The Failure of the American Legal Response to the Computer' (1992) 43 Hastings Law Journal 1321, 1348.

31 ibid.

32 (2017) 251 3d 1168 (Southern District of Texas).

33 ibid.

as 'poor performers' by the software.[34] The court held that the teachers' procedural due process rights were violated as the algorithms were proprietary information and could not be effectively scrutinized and challenged by the teachers.[35] This case demonstrates the necessity of an explanation for how the AI software deemed certain teachers as 'poor performers', particularly as the results of the software ultimately led to the decision to terminate certain teachers' employment.

In *Wisconsin v Loomis*,[36] the Wisconsin Supreme Court held a trial court's use of the COMPAS tool in its decision to sentence Mr Loomis, which did not violate his due process rights despite the fact that the methodology and algorithms used by the tool were not disclosed to the court or Mr Loomis.[37] The court explained that the COMPAS tool was not the sole basis for Mr Loomis' sentencing and that the court ultimately retained the discretion to disagree with any COMPAS assessments when appropriate.[38] However, it is difficult to understand how and when judges will decide to rely on automated risk assessments such as COMPAS, when the precise methodology of these automated systems is not transparent. These cases demonstrate the need for transparency and the ability for individuals to challenge and review automated decisions, particularly when they significantly impact upon their dignity and liberty.[39]

A THEORETICAL JUSTIFICATION FOR A RIGHT TO EXPLANATION FOR AUTOMATED DECISIONS

Information privacy and control over automated decision-making

It is widely agreed that privacy is central to an individual's creation and expression of their identity which is in turn dependent upon withholding and disclosing aspects of their self.[40] Under such a conception of privacy, 'information privacy' emerged in response to the development of information technology in the 1960s and 1970s.[41] Information technology propelled the widespread use of computers to automate the processing of personal information for administration purposes.[42] Information privacy rights thus emerged to support individual autonomy by providing control over the access, use and disclosure of their personal information.[43] In a digital world where almost every online interaction is now used as a new data point and combined

34 ibid.

35 ibid.

36 (n 20) 760.

37 ibid.

38 ibid.

39 See Zalnieriute, Moses and Williams (n 12) 443.

40 David Lindsay, 'An Exploration of the Conceptual Basis of Privacy and the Implications for the Future of Australian Privacy Law' (2005) 29 Melbourne University Law Review 131, 169.

41 See also David Banisar and Simon Davies, 'Global Trends in Privacy Protection: An International Survey of Privacy, Data Protection, and Surveillance Laws and Developments' (1999) 18 John Marshall Journal of Computer and Information Law 1, 10.

42 ibid.

43 See for example Brian Fitzgerald and others, *Internet and E-Commerce Law* (Thomson Reuters 2011) 898; Moira Paterson, 'HealthConnect and Privacy: A Policy Conundrum' (2004) 12 Journal of Law and Medicine 80, 81; Charles Fried, Privacy (1968) 77 Yale Law Journal 475, 482.

with other data points to inform automated decision-making systems,[44] it is necessary for individuals to retain control over the use of their personal data.

The provision of transparency rights such as a right to explanation is inextricably linked to the ideal of effective control over automated decision-making and has been extensively considered in existing legal and computational science literature.[45] Burrell has argued that explanations refer to attempts to convey the internal state or logic of machine learning algorithms that produce automated decisions.[46] The dominant position in the legal literature justifies the requirement of transparency through explanations on the basis that an individual adversely affected by an automated decision has the right to 'understand why' and frames this in deontological terms of control and dignity as a human being.[47] However, there is now an emerging scholarly debate about how explanations should work in practice.[48] The consensus in the legal literature on what level of transparency is needed, whether different types of industries require different solutions and whether explanations require the full disclosure of proprietary algorithms, is lacking clarity.[49]

Deontological and consequentialist approaches to privacy protection

Historically, there have been two theoretical approaches that have applied to determine the extent and shape of privacy protection, namely deontological and consequentialist approaches. Delineating between these two approaches to privacy protection is helpful to develop a theoretical justification for laws to address the transparency issues posed by automated decisions. For deontologists, the value of privacy is inherently rights-based and is determined by the extent to which it is consistent with basic moral rights and duties,[50] such as promoting individual autonomy, dignity and control.[51] Deontologists ground their arguments in Kantian ethics which state that at the core of dignity is the notion that each individual should be treated as an end in him or her, rather than as a means to furthering another person's or society's ends.[52] In marked contrast, consequentialists determine the extent to which privacy's legal protection promotes maximum desirable goals, such as maximizing efficiency or economic rationality.[53]

44 See for example Patrick Lowden and Andrew Booth, 'Unlocking the Potential of Data in Australia's Financial System' (2019) 29 Journal of Banking and Finance Law and Practice 332, 338.

45 See Lilian Edwards and Michael Veale, 'Slave to the Algorithm? Why a 'Right to an Explanation' is Probably Not the Remedy You Are Looking For' (2017) 16 Duke Law and Technology Review 18, 41; Pasquale (n 3) 3; Joshua Kroll and others, 'Accountable Algorithms' (2017) 165 University of Pennsylvania Law Review 633, 658.

46 Burrell (n 5) 10.

47 Tal Zarsky, 'Transparency in Data Mining: From Theory to Practice' in Bart Custers et al (eds), *Discrimination and Privacy in the Information Society: Data Mining and Profiling in Large Databases* (Springer 2013) 301, 310.

48 See Kroll and others (n 45) 658; Deven Desai and Joshua Kroll, 'Trust But Verify: A Guide to Algorithms and the Law' (2017) 31 Harvard Journal of Law & Technology 2, 4.

49 Desai and Kroll, ibid.

50 Lindsay (n 40) 144; Neil McCormick, *Legal Right and Social Democracy* (OUP 1982) 144.

51 Lindsay, ibid 134.

52 Stanley I Benn, 'Privacy, Freedom, and Respect for Persons' in Roland Pennock and John Chapman (eds), *Privacy* (Atherton Press 1971) 1, 16–26.

53 ibid.

Differing arguments have been advanced to support such a deontological basis for privacy protection. Fried likened the concept of privacy to a 'principle of morality' which values 'the equal liberty of each person to define and pursue his or her values free from undesired impingements by others'.[54] According to this view, an individual is entitled to the respect of others to freely pursue his or her values.[55] Respect is directly linked with an individual's ability to enter into intimate relationships of love, friendship and trust, which in turn, depends upon the individual's ability to control and reveal selected private information to intimate others.[56] For Fried, privacy is therefore valued because it is a prerequisite for the formation of intimate relations and essential to what it means to be a person.[57]

Reiman, however, provides a more nuanced deontological justification for privacy protection.[58] In contrast to Fried, Reiman posits that an individual's claim to privacy is distinguished from their desire to pursue intimate relations.[59] While Reiman agrees with Fried that privacy is linked to one's moral personhood, he departs to argue that privacy is essential to the social construction of personal identity because it enables an individual to experience moral ownership of his or her existence.[60] Reiman's conception of privacy is persuasive as it addresses the critical flaw inherent in Fried's theory that those incapable of forming intimate relationships cannot claim a right to privacy.[61]

In contrast, consequentialists value privacy by the extent to which privacy protection promotes desirable results, such as maximizing efficiency or social welfare.[62] Consequentialist justifications are based on utilitarianism which states that maximum happiness is the fundamental goal, and rights are only valued derivatively for their instrumental capacity to serve maximum happiness.[63] These justifications are achieved by first determining what is 'good' and then defining 'the right' that will promote 'the good'.[64] For example, if 'the good' were to achieve greater efficiency in automated decision-making systems, then a consequentialist approach would only consider the legal protection of privacy and personal data to the extent that it maximizes efficiency.

Deontological scholars argue that consequentialist approaches to privacy protection will detrimentally privilege broader social goals at the cost of individual autonomy, control and dignity. Rawls constructs a persuasive argument against consequentialism and asserts that consequentialists do not take seriously the distinction between individuals.[65] The primary issue with consequentialism is that

54 Fried (n 43) 479.
55 ibid.
56 ibid 480–82.
57 ibid.
58 Jeffrey Reiman, 'Privacy, Intimacy, and Personhood' in Ferdinand Schoeman (ed), *Philosophical Dimensions of Privacy: An Anthology* (CUP 1984) 300, 310.
59 ibid.
60 ibid.
61 ibid.
62 McCormick (n 50) 144.
63 See John Rawls, *A Theory of Justice* (Harvard UP 1971) 24–30.
64 ibid.
65 ibid 27.

impersonal social aims will always outweigh the interests and values of individuals, despite it not being rational for the utility of one person to be sacrificed for a greater increase in the utility of others.[66] Information privacy laws that are grounded in consequentialist aims will prefer to improve technocratic procedures of information management and the efficient processing of personal data, rather than addressing the concerns and implications presented by opaque automated decisions.[67]

From a consequentialist perspective, the collection and processing of large amounts of personal data by automated decision-making systems is a deeply impersonal form of social practice which entails treating individuals as merely data subjects in instrumental terms.[68] These processes severely interfere with the deontological aims of preserving legitimate individual rights and interests.[69] This article therefore argues that a deontological justification to information privacy regulation is preferred as it aims to address the transparency issues posed by automated decision-making systems by privileging individual autonomy, control and dignity.[70]

DETERMINING THE APPROPRIATE REQUIRED LEVEL OF TRANSPARENCY AND ACCOUNTABILITY

If it is accepted that a right to explanation is necessary to address the lack of transparency and loss of control issues presented by automated decision-making systems, the next step is to consider the nature and ambit of such a right. Should such a right be framed to impose total transparency on automated decision-making systems or should it be limited in some fashion? To what extent can counterfactual explanations address the concerns raised by commercial entities over the disclosure of their proprietary algorithms?[71] This section will critically analyse the proposition that total transparency is a prerequisite to the effective regulation of automated decision-making systems and progress to considering the merits of counterfactual explanations as a viable alternative to total transparency.[72]

Total transparency in automated decision-making

'Explanations' refer to an attempt to convey the internal state or logic of an algorithm that produces an automated decision.[73] Some commentators have argued that the right to explanation requires entities to disclose their proprietary algorithms to individuals subject to automated decisions.[74] While this article maintains that transparency rights are an effective control mechanism over automated decision-making systems,[75] it departs to argue that it is unnecessary and

66 Lindsay (n 40) 152.
67 ibid 160.
68 ibid 163.
69 ibid.
70 See Julie Cohen, 'What Privacy is For' (2013) 126 Harvard Law Review 1904, 1907.
71 Sonia Katyal, 'Private Accountability in the Age of Artificial Intelligence' 66 University of California Los Angeles Law Review 54, 56.
72 Wachter, Mittelstadt and Russell (n 4) 843.
73 Burrell (n 5) 10.
74 Compare Kroll and others (n 45) 657; Wachter, Mittelstadt and Russell (n 4) 843.
75 Edwards and Veale (n 45) 41.

undesirable for entities to disclose their proprietary algorithms.[76] Kroll and others provide a compelling argument against total transparency as he identified that individuals may manipulate and undermine automated decision-making systems when they are provided their algorithms.[77] For example, if an automated credit rating system discloses its credit risk behavioural indicators, then individuals may try to avoid being linked to those indicators while still engaging in risk-generating behaviour.[78] With this in mind, total transparency with the full disclosure of an automated decision-making system's rules and criteria has the potential to subvert its efficiency and fairness.[79]

Total transparency of an automated decision-making system is unnecessary due to the inherent human limitations of truly understanding or explaining the operation of machine learning algorithms used by second wave decision-making systems.[80] Burrell has identified the difficulty for technical experts to understand the results of certain machine learning algorithms, particularly when they are programmed to analyse large volumes of data.[81] For example, an automated decision-making system employing facial recognition involves 'a complex combination of distal relationships, angles, colouring, and shape, combined through a multi-layered neural network'.[82] If a right to explanation requires total transparency, then this would place onerous and impractical obligations on entities to interpret and explain automated decisions involving complex machine learning algorithms for each individual case.[83]

In a similar vein, even if the rules or the actual machine learning algorithms were disclosed, individuals without the requisite technical knowledge would not be able to extract a meaningful explanation from such disclosure.[84] Individuals would require a technical expert or the original programmer of the automated system.[85] In circumstances where the implications of an automated decision are significant, the inability to access a technical expert to extract a rationale for the decision effectively reduces the extent to which the decision itself can be described as transparent.[86] As technology develops and machine learning algorithms become more complex over time, it will become less likely that human comprehensibility of the internal logic of machine learning algorithms can be achieved.[87] It is thus unnecessary and undesirable for a right to explanation to impose total transparency on automated decision-making systems.

76 Kroll and others (n 45) 654.
77 ibid.
78 ibid.
79 ibid.
80 Burrell (n 5) 10.
81 ibid.
82 Zalnieriute, Moses and Williams (n 12) 442.
83 See John Zerilli and others, 'Transparency in Algorithmic and Human Decision-Making: Is there a Double Standard' (2019) 32 Philosophy & Technology 661–683.
84 Burrell (n 5) 10.
85 ibid.
86 ibid.
87 Zalnieriute, Moses and Williams (n 12) 442.

The value of counterfactual explanations

In contrast to explanations that involve an attempt to outline the logic of algorithms,[88] counterfactual explanations specify which circumstances would need to change to achieve a more desirable decision.[89] If an individual's loan application, for example, were rejected by an automated system, a counterfactual explanation would provide the individual an assessment of the minimum change needed for the application to be successful.[90] In this example, the counterfactual explanation would state that the loan would have been granted if the applicant applied for a loan of $10,000 or less if their income was $25,000 or more. In this way, counterfactual explanations address the concerns raised by commercial entities over the disclosure of their proprietary algorithms and trade secrets with transparency rights as such explanations are provided in a way that produces a minimal amount of information capable of altering a decision.[91]

Counterfactual explanations attempt to address the human interpretability issues inherent in machine learning algorithms.[92] Counterfactual explanations do not require individuals to understand any algorithms in order to extract a meaningful explanation. They are easy to understand and practically useful as they provide the circumstances that need to change to achieve a more desirable decision.[93] In contrast, the disclosure of algorithms to individuals would not provide individuals any value as they are difficult to render comprehensible for non-experts. Therefore, counterfactual explanations are a viable alternative to total transparency as they balance the commercial interests of private entities and the legitimate rights and interests of individuals.[94]

PRESENT AUSTRALIAN LAW AND POLICY

Amending laws to address technological disruption

As technological innovation and disruption is an accelerating feature of modern society, it is important to devise a formal model for how the law should respond to such constant change. Historically, legal responses to technological change were often premised on a belief that technological innovations were one-off events that could be addressed by discrete and specific reform.[95] Such an assumption of the 'arc' effect of technological change is apparent in Spar's articulation of the four stages of technological growth.[96] Spar notes that the initial development and discovery of an innovative technology is followed by a second phase of commercialism, followed by a third phase of growing conflict between the dictates of ordinary commerce and the dynamic and unrestrained spirit of the new technology, termed creative anarchy,

88 Burrell (n 5) 10.
89 Wachter, Mittelstadt and Russell (n 4) 843.
90 ibid.
91 Katyal (n 71) 56.
92 Wachter, Mittelstadt and Russell (n 4) 843.
93 ibid.
94 ibid 844.
95 See Debora Spar, *Ruling the Waves: Cycles of Discovery, Chaos, and Wealth from the Compass to the Internet* (Harcourt 2001).
96 ibid.

followed by the final phase of lawmakers formulating rules to regulate this technology.[97] This model embodies an implicit assumption that there will be a stage when the 'flashpoint of discovery' will dim and enable the enacted laws to effectively govern for the foreseeable future.[98] However, it is relevant to note that Spar's model is based on early experiences of widespread technological change, such as the development of Atlantic trade routes, telegraph and broadcasting.[99] In marked contrast, the development of digital technologies has maintained a constantly upward trajectory of innovation and disruption. Given the continual accelerating change in modern society, such an approach can lead to regulatory gaps as well as inconsistent or overlapping laws that fail to regulate similar matters in a comparative and holistic way.[100]

In such a context, this article proposes that the following three-step model could be useful in guiding law and policy responses to technological change such as the increasing reliance on automated decisions.[101] When confronted with a technological disruption, it is first necessary to analyse its effects across the full spectrum of laws. As technological change does not respect the established boundaries of legal scholarship, such as contract, torts or administrative law, and typically elicits a variety of interconnected and complex effects, such a holistic approach supports both the avoidance of double-regulation and regulatory gaps.[102] As this first stage commonly involves understanding a variety of interconnected statutory frameworks, it will require an interdisciplinary approach, integrating technical expertise from outside of the field of law. Building on this understanding, it is secondly necessary to consider whether and to what extent these identified existing laws could apply to the technological disruption. This is a legal exercise of statutory interpretation that considers the nature and purpose of laws and their likely ambit of operation.[103] If existing laws cannot be effectively extended to govern the new technology, the third and final stage is to determine whether existing laws can be changed to address the technological change or whether it is necessary to design a new legal framework. In drafting such laws, whether it be specific reform, changes to an existing matrix of laws or the design of a whole new legal framework, it is important to adopt technology-neutral principles of drafting to support the longevity of the new laws. Such an approach will ensure laws remain flexible and adaptable for future technological evolution.

Applying the above model to the present case of automated decision-making, the following section of this article will examine present Australian privacy and administrative laws in order to identify whether these laws confer a right to receive an explanation for automated decisions. This analysis will then be followed by a consideration of the ambit of any such identified laws and an examination of whether

97 ibid.

98 ibid.

99 ibid.

100 Kevin Werback, 'Breaking the Ice: Rethinking Telecommunications Law for the Digital Age' (2005-2006) 4 Journal on Telecommunications & High Technology Law 59, 71–74.

101 Niloufer Selvadurai, 'The Relevance of Technology Neutrality to the Design of Laws to Criminalise Cyberbullying,' (2018) 1 International Journal of Law and Public Administration 1.

102 ibid.

103 Winston Maxwell, 'Technology Neutrality in Internet, Telecoms and Data Protection Regulation (2014) 31 Computer and Telecommunications Law Review 1.

they can be expansively interpreted to confer a right to explanation for automated decisions.

The applicability of existing Australian statutory frameworks

It is useful to consider whether and to what extent existing Australian statutes governing transparency and accountability in the use and application of personal data could apply to regulate automated decisions. While the Privacy Act 1988 (Cth)[104] does not create an express right to explanation for decisions relating to an individual's data, it does impose certain obligations on entities managing data that may potentially extend to creating a right to explanation for automated decisions. The Privacy Act seeks to balance the protection of the privacy of individuals with the interests of entities in carrying out their functions or activities.[105] The Privacy Act also aims to promote the responsible and transparent handling of personal information by entities and provide a means for individuals to complain about an alleged interference with their privacy.[106] The Privacy Act regulates 'APP entities', individuals, body corporates, partnerships, unincorporated associations or trusts that have an annual turnover of $3,000,000 or more for a financial year.[107] Parties who have an annual turnover of less than $3,000,000 a financial year are also subject to the Privacy Act if they provide a health service or hold health information other than in an employee record.[108] Furthermore, the 'Australian Privacy Principles' (APPs) form a Schedule to the Privacy Act and articulate standards, rights and obligations relating to the collection, use and disclosure of personal information to strengthen good governance and accountability.[109]

Most relevant to the present discussion is APP 1 which obliges APP entities to be open and transparent in the way they manage personal information.[110] APP entities are also required to take such steps as are reasonable in the circumstances to implement 'practices, procedures and systems' relating to the entity's 'functions or activities' to ensure the privacy of personal information they collect, use and store.[111] In this regard, APP 1.2(b) further stipulates that such 'practices, procedures and systems' must be sufficient to 'enable the entity to deal with inquiries or complaints from individuals about the entity's compliance with the APP'.[112] Pursuant to APP 1, APP entities must have a privacy policy that provides specific information on such matters as the types of personal information it collects and the procedure by which an individual may complain about any suspected breach of the APPs.[113] The 'Explanatory Memorandum' to the Privacy Amendment (Enhancing Privacy Protection) Bill 2012 notes that such practices could include designing and

104 (hereafter 'Privacy Act').
105 ibid, pt 1, s 2A(b).
106 ibid, pt I, sect 2A(d).
107 ibid, pt III, div II, s 15.
108 ibid, pt II, div I, s 4(b).
109 Privacy Amendment (Enhancing Privacy Protection) Act 2012 (Cth) sch 1 (hereinafter 'APP').
110 ibid 1.1.
111 ibid 1.2(a).
112 ibid 1.3.
113 ibid 1.2(b).

implementing 'systems or infrastructure' for the collection and handling of personal information.[114] In the present case, this provision could be viewed as supporting an individual's inquiry as to the basis of the algorithm used to generate an automated decision. This article submits that such an expansive interpretation of APP 1 would enable it to support a limited right to obtain an explanation for automated decisions in the context of an individual's personal privacy.

However, as intimated above, the right to explanation that could conceivably be supported by the Privacy Act would be extremely narrow. This is because any such right to explanation would be limited to automated decisions that relate to the protection of the privacy of individuals. Hence, the present protection provided by the Privacy Act would not apply where an individual desires an explanation for an automated decision that does not relate to an infringement of privacy but rather relates to another legitimate interest, such as equitable treatment in the assessment of a loan application or a social welfare application. Moreover, the limited scope of the Privacy Act's application to entities with an annual turnover of $3,000,000 or more for a financial year unless the entity's role relates to health, significantly limits its ambit. Thus, the present protection provided by the Privacy Act does not sufficiently encompass the diverse use cases of automated decision-making and the varied impacts on the lives of individuals.

While privacy law is the primary prism through which this article has analysed the creation of a right to explanation for automated decisions, it is also relevant to examine present Australian administrative law. The Federal government's *Automated Assistance in Administrative Decision-Making Better Practice Guide* (*Guide*) recognizes that automated systems can play a significant and beneficial role in administrative decision-making but stipulates that administrative authorities who use computer systems for administrative decision-making should adhere to administrative law values when developing and operating such systems.[115] To this end, the *Guide* provides a checklist to ensure the appropriate management of decisions through the full life cycle of an automated system. However, the recent decision of the Full Federal Court in *Pintarich v Deputy Commissioner of Taxation* (*Pintarich*)[116] suggests that Australian administrative law does not view automated decisions as being administrative decisions. In *Pintarich*, a majority of the Full Federal Court held that 'no decision was made unless, accompanied by the requisite mental process of an authorised officer'.[117] Hence, the rights that an individual ordinarily has to be provided, an explanation for an administrative decision, as well as the requirement for such an administrative decision to be reasonable, are unlikely to attach to automated decisions.

While a variety of Australian statutes do permit computer-assisted decision-making,[118] none of these statutes confer a right to explanation for such automated or semi-automated decisions. For example, section 6A of the Social Security

114 Explanatory Memorandum, Privacy Amendment (Enhancing Privacy Protection) Bill 2012 (Cth) 79.

115 Australian Government, *Automated Assistance in Administrative Decision-Making Better Practice Guide* (Department of Industry, Science, Energy and Resources, Australian Government, Canberra, 2007).

116 [2018] FCAFC 79.

117 ibid.

118 Kobi Leins, 'What is the Law When AI Makes the Decisions?' (*The University of Melbourne*, 4 December 2019) <https://pursuit.unimelb.edu.au/articles/what-is-the-law-when-ai-makes-the-decisions> accessed 15 January 2020.

(Administration) Act 1999 (Cth) provides that the Secretary may 'arrange for use of computer programs to make decisions ... under the Secretary's control ... for any purposes for which the Secretary may make decisions under the social security law', and does not confer a right to explanation as to any algorithm used to generate such a decision. Further, while the A New Tax System (Family Assistance) (Administration) Act 1999 (Cth) similarly provides that the Secretary may 'arrange for use of computer programs to make decisions ... under the Secretary's control ... for any purposes for which the Secretary may make decisions under the family assistance law', the Act also does not confer any such right to explanation. In the field of finance and superannuation, the National Consumer Credit Protection Act 2009 (Cth), section 242, contains an analogous provision authoring the Australian Securities and Investment Commission to use computer programs to make decisions under the National Consumer Credit Protection Act and the Superannuation (Government Co-contribution for Low Income Earners) Act 2003 (Cth), section 48, permits the Commissioner to make such computer-assisted decisions in relation to this Act and its regulation. In the field of education, such computer-assisted decision-making is also allowed under the Australian Education Act 2013 (Cth) and the VET Student Loans Act 2016 (Cth), section 105. In the health sector, the National Health Act 1953 (Cth), section 101B and the Aged Care Act 1997 (Cth), section 23B-4, permits such decisions. However, none of the above statutes or other statutes that enable automated decision-making[119] are accompanied by a statutory right to explanation as to the rationale or basis for the automated component of the decision.

As present Australian law does not confer a right to explanation for automated or semi-automated decisions, it is useful to examine the law reform discourse to date and the operation of the European Union's GDPR in order to determine whether further reform is required in this area in Australia.

The Australian law reform discourse

The governance of AI work has been the subject of intense reform debate and discussion in Australia. In 2019, the Federal Government released its 'AI Ethics Framework',[120] comprising eight principles that should govern the design, development, integration and use of AI systems. The stated objective of the 'AI Ethics Framework' is to support better outcomes, reduce negative impact and apply the highest standards of ethical business and good governance to AI. While the 'AI Ethics Framework' is voluntary and intended to complement rather than replace legislation, the principles provide an ethical framework for the use of AI. The 'AI Ethics Framework' stipulates that governing principles of AI processes and systems should be human, social and environmental well-being,[121] human-centred values,[122]

119 The following table lists the statutes which contemplate and enable automated decision-making: <https://airtable.com/shrpkHgfDpvec6BA3/tblHPWVuiNI6v63nn?backgroundColor= blue&blocks= hide>.

120 Australian Government, 'AI and Ethics Framework' (2019) <https://www.industry.gov.au/data-and-publications/building-australias-artificial-intelligence-capability/ai-ethics-framework> accessed 7 January 2020.

121 Principle 1.

122 Principle 2.

fairness,[123] privacy protection and security,[124] reliability and safety,[125] transparency and explainability,[126] contestability[127] and accountability.[128]

Of particular relevance to this article are Principles 6, 7 and 8 and the relationship between them. Principle 6 stipulates the need for transparency and responsible disclosure to enable individuals to know when they are being 'significantly impacted by an AI system' and information on enabling such individuals to find out 'when an AI system is engaging with them'.[129] Contestability is supported by Principle 7 that stipulates that where an AI system 'significantly impacts a person' they should be afforded a timely process to 'challenge the use or output' of such an AI system.[130] Fortifying these principles is that of accountability which provides that 'those responsible for the different phases of the AI system lifecycle' should be 'identifiable and accountable for the outcomes of the AI systems'.[131] Principle 8 also contains a passing reference to the need to enable 'human oversight of AI systems'.[132] More broadly relevant is Principle 2 on human-centred values which stipulates that AI systems 'should respect ... the autonomy of individuals',[133] complemented by Principle 1 which pronounces that AI systems should benefit individuals and society.[134]

However, while the Australian 'AI Ethics Framework' principles are useful in articulating the overarching fundamental values that should govern AI processes and systems, they do not provide guidance in designing substantive legal provisions to uphold these values. For example, the reference to 'human oversight of AI systems' in Principle 8 does not clarify the nature or extent of such oversight.[135] Moreover, Principle 8 fails to address the area of uncertainty created by *Pintarich* as to the degree of human involvement or oversight that is required for the final decision to qualify as an administrative decision.[136] Thus, considerable additional work is needed to translate these broad principles into a specific statutory provision that confers a right to explanation.

The 2019 'OECD Recommendations on Artificial Intelligence' (OECD Recommendations)[137] extends the Australian 'AI Ethics Framework' by providing a more fine-grain governance framework. Similar to the Australian 'AI Ethics Framework', the OECD Recommendations seek to promote the responsible stewardship of trustworthy AI while also upholding respect for fundamental human rights

123 Principle 3.
124 Principle 4.
125 Principle 5.
126 Principle 6.
127 Principle 7.
128 Principle 8.
129 Principle 6.
130 Principle 7.
131 Principle 8.
132 Principle 8.
133 Principle 2.
134 Principle 1.
135 Principle 8.
136 Principle 8.
137 OECD, adopted by the OECD Council at Ministerial level on 22 May 2019 on the proposal of the Committee on Digital Economy Policy (CDEP). Adopted on 22 May 2019; In June 2019, G20 Leaders welcomed G20 AI Principles, drawn from the OECD Recommendation.

and democratic values. To this end, the OECD Recommendations articulate five values-based principles, being inclusive growth, sustainable development and well-being, human-centred values and fairness, transparency and explainability, robustness, security and safety, and accountability.[138] However, unlike the Australian guidelines, the OECD Recommendations also provide detailed definitions of the parties, systems and stakeholders to be governed. In this respect, 'AI system' is defined as 'a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments' which is 'designed to operate with varying levels of autonomy'. Unlike the Australian 'AI Ethics Framework' which frames principles to apply to the full AI 'life cycle' but does not identify the stages of such a life cycle, the OECD articulates the full schedule of AI use. The 'AI system life cycle' is defined to encompass 'design, data and models', as well as data collection and processing, followed by a model building stage that involves 'verification and validation'. This is followed by a third 'deployment' stage and a final 'operation and monitoring' phase. The OECD Recommendations acknowledge that these phases often occur in an iterative manner and are not always sequential.

The fine-grain formulations of the OECD Recommendations are also apparent in its differentiation of 'AI systems' from 'AI knowledge'. While common in technical specifications, such sensitive delineations are not commonly seen in legal principles and laws. 'AI knowledge' is defined as being 'the skills and resources, such as data, code, algorithms, models, research, know-how, training programmes, governance, processes and best practices, required to understand and participate in the AI system lifecycle'. Finally, 'AI actors' are termed to be those 'who play an active role in the AI system lifecycle, including organisations and individuals that deploy or operate AI'. Importantly, the reference to 'active role' in the definition of AI actors recognizes that AI-generated outcomes can have varying levels of machine and human interaction. Rather than articulating a degree of involvement, such as 'substantial role' or 'significant role', the OECD Recommendations have opted for 'active' contribution, suggesting that causative contribution than total overall level of contribution will be the deciding criteria in determining whether a party is an AI actor. Such a sensitive formulation recognizes the varying degrees of human intervention and oversight that are encompassed within automated decision-making systems.

Most significantly for the present discussion and in marked contrast to the Australian 'AI Ethics Framework', the OECD Recommendations expressly refers to a right to explanation. Clause 1.3 notes the need for AI actors to commit to transparency and responsible disclosure regarding AI systems and outlines the processes to be followed to achieve this goal.[139] The clause notes the obligation for AI actors to provide 'meaningful information, appropriate to the context, and consistent with the state of art' to foster 'a general understanding of AI systems and make stakeholders aware of their interactions with AI systems . . . enable those affected by an AI system to understand the outcome'.[140] Furthermore, unlike the Australian 'AI Ethics Framework', the OECD Recommendations expressly refer to a right to explanation.

138 G20 Ministerial Statement on Trade and Digital Economy Annex 1, Cl 1.

139 ibid.

140 ibid.

Clause 1.3(iv) stipulates that such 'meaningful information' must enable those adversely affected by an AI system to challenge its outcome based on 'the logic that served as the basis for the prediction, recommendation or decision'.[141] This provision is not limited in its sphere of application and imposes an additional layer of obligation on top of that already provided by existing statutes.

While the Australian 'AI Ethics Framework' clearly supports transparency and accountability in the use of AI, its principles are broadly framed and do not precisely formulate the duties and obligations of AI actors. Moreover, in contrast to the OECD AI Recommendations, it does not expressly address the issue of a right to explanation for automated decisions. In the absence of developed Australian law or policy in this area, the next chapter of this article will consider the European Union's GDPR in order to consider a reform agenda for Australia.

ENACTING A NEW STATUTORY RIGHT TO EXPLANATION FOR AUTOMATED DECISIONS

Insights provided by the GDPR's right to explanation

In the absence of any Australian right to explanation for automated decisions, the right to explanation in the European Union's GDPR provides valuable insights. While the legal status of the relevant provision has been the subject of considerable debate,[142] it is useful to examine the provision in detail, including its theoretical basis, ambit of operation and effectiveness in giving statutory effect to the above-discussed principles of transparency and accountability. The GDPR governs the relationships of data subjects and data controllers.[143] The former are defined by Chapter 5 to be persons whose personal data is being collected, held or processed, while the latter encompass public authorities, agencies or any other body which alone or jointly with bodies determines the purposes and means of the processing of personal data.[144]

Articles 13, 14 and 22 of the GDPR create a variety of rights and requirements relating to the collection of information from individuals and automated decision-making. Article 131 stipulates that where a data controller collects personal data from a data subject, the data subject should have the right to obtain certain information such as the purpose of the processing and the recipients of the data.[145] Additionally, Article 13.2 imposes certain additional obligations to ensure 'fair and transparent processing'.[146] This includes 'meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject' pursuant to Article 13.2(f).[147] Article 14 creates a similar

141 ibid 1.3(iv).

142 Margot Kaminski, 'The Right to Explanation, Explained' (2018) 34 Berkeley Technology Law Journal 189.

143 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L 119/1; cor. OJ L 127, 23.5.2018.

144 ibid, ch 5.

145 ibid, art 13.1.

146 ibid, art 13.2.

147 ibid, art 13.2(f).

obligation where the data have not been collected from the data subject, with Article 14.2(g) creating an analogous obligation to that created by Article 13.2(f).[148] Article 22.1 stipulates that a data subject should have 'the right not to be subject to a decision based solely on automated processing ... which produces legal effects concerning him or her or similarly significantly affects him or her'.[149] An exemption applies where the data subject has consented to the decision or where the decision is authorized by a national law which also creates suitable measures to 'safeguard the data subject's rights and freedoms and legitimate interests'.[150]

While not forming part of the text of the GDPR, Recital 71 provides guidance on the interpretation of Articles 13, 14 and 22. Recital 71 stipulates that data subjects should have 'the right not to be subject to a decision, which may include a measure, evaluating personal aspects relating to him or her which is based solely on automated processing and which produces legal effects concerning him or her or similarly significantly affects him or her'.[151] An exemption for decision-making based on such processing is created in circumstances where it is expressly authorized by Union or Member State law to which the controller is subject. However, even when the exemption applies, the data subject must have the right to be protected by 'suitable safeguards', including specific information and the right to 'obtain an explanation of the decision reached'.[152] While the GDPR does not specify what constitutes 'legal effects concerning him or her or similarly significantly affects him or her', Recital 71 provides certain examples of decisions with significant legal effect.[153] The examples provided include the automatic refusal of an online credit application and e-recruiting practices without any human intervention.[154] The *Working Party Guidelines* further clarifies that 'only serious impactful effects' will be encompassed by Article 22[155] and provides additional examples of decisions that affect financial circumstances, access to health services, access to education, deny employment or put someone 'at a serious disadvantage'.[156]

Critical to the present discussion, Recital 71 states that 'such processing should be subject to suitable safeguards, which should include ... the right to obtain human intervention, to express his or her point of view, to obtain an explanation of the decision reached after such assessment and to challenge the decision'.[157] Edwards and Veale and Wachter and others suggest that this provision does not create a right to explanation because it is not within the text of the GDPR.[158] However, Kaminski

148 ibid, art 14.
149 ibid, art 22.1.
150 ibid, art 22.2.
151 ibid, rec 71.
152 ibid, art 22(3).
153 ibid, rec 71.
154 ibid.
155 Article 29 Data Protection Working Party, Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679,17/EN (6 February 2018) 21.
156 ibid.
157 GDPR, rec 71.
158 Edwards and Veale (n 45) 50 ('Our view is that these certainly seem shaky foundations on which to build a harmonized cross-EU right to algorithmic explanation'); Sandra Wachter, Luciano Floridi and Brent Mittelstadt, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist' (2017) 7 International Data Privacy Law 76, 79.

argues that such a position is incorrect. She argues that the right to challenge a particular decision, which is in the text of the GDPR, is contingent upon the ability to obtain relevant information as to the basis of the decision, stating 'an individual has a right to explanation of an individual decision because that explanation is necessary for her to invoke the other rights' such as the right 'to contest a decision, to express her view-that are explicitly enumerated in the text of the GDPR'.[159] Kaminski's argument is largely consistent with those expressed by Mendoza and Bygrave[160] and Selbst and Powles.[161]

Irrespective of whether such a right to explanation is binding in European Union law, the GDPR's framing of the right provides a useful template for consideration. A right to explanation is triggered in circumstances where:

a. the decision was solely based on automated processing;
b. the decision produced legal effects concerning the individual; and
c. these effects were significant.

If these three criteria are established, the organization is obliged to provide 'meaningful information about the logic involved' in making the decision.

In the context of the lack of transparency in automated decisions discussed in Chapter I, it is suggested that such a right to explanation is critical to give effect to an individual's right to challenge decisions that affect him or her. The corollary of this is that an absence of such a right to explanation in the age of AI serves to substantially undermine the effectiveness of a right to review or challenge decisions. Thus, the purpose of the next section of this article is to consider how such a right to explanation for automated decisions could potentially be enacted in Australia.

Enacting a statutory right to explanation in Australia

Prior to determining whether and to what extent Australia should be guided by the GDPR in enacting a right to explanation, it is useful to compare the philosophical underpinnings of European Union and Australian laws in this area in order to determine the feasibility of such a reform course. The European Union's privacy rights are premised on the need to protect and respect personal autonomy and dignity.[162] Reidenberg describes European governance philosophy to be a:

> vision of governance generally regards the state as the necessary player to frame the social community in which individuals develop, and information

159 Kaminski (n 142) 204.

160 Isak Mendoza and Lee Bygrave, 'The Right Not to be Subject to Automated Decisions Based on Profiling' in Eleni Synodinou and others (eds), *EU Internet Law: Regulation and Enforcement* (Springer 2017) 16.

161 Andrew Selbst and Julia Powles, 'Meaningful Information and the Right to Explanation' (2017) 7 International Data Privacy Law 233, 242.

162 James Whitman, 'The Two Western Cultures of Privacy: Dignity Versus Liberty' (2004) 113 Yale Law Journal 1151, 1161.

practices must serve individual identity. Citizen autonomy, in this view, effectively depends on a backdrop of legal rights.[163]

According to Reidenberg, rights-based legal protection of privacy is necessary for the development of a desirable society where individuals are empowered to pursue autonomy and individual identities.[164] Reflective of such thinking, the GDPR's right to explanation is inherently grounded in a deontological justification for information privacy protection as it promotes individual autonomy, dignity and control over automated decisions. The right to explanation provides individuals an explanation for how an automated decision was made, despite its impact on the efficiency and consistency of data processing in automated decision-making systems. This ultimately addresses the loss of dignity created by opaque automated decisions as individuals attain an understanding of how decisions about them were made.[165] In a similar vein, a right to explanation provides individuals greater control over automated decision-making systems as it may allow individuals to ensure the accurate processing of their personal data.[166]

In contrast to the European Union, Australia has historically adopted a consequentialist approach to information privacy regulation, as evidenced by its departure from rights-based approaches to protecting human rights and privacy.[167] Article 12 of the Universal Declaration of Human Rights (UDHR) and Article 17 of the International Covenant on Civil and Political Rights (ICCPR) both provide that 'no one shall be subjected to arbitrary or unlawful interference with his privacy, family, home or correspondence, nor to unlawful attacks on his honour and reputation.'[168] The UDHR and ICCPR prompted judicial development in New Zealand, the UK and Canada[169] which all now recognize a right to privacy.[170] However, Australia has refrained from introducing a Bill of Rights which would implement Article 12 of the UDHR and Article 17 of the ICCPR, and has refused to become a party to any binding international human rights instruments.[171] Australia's traditional consequentialist approach to information privacy regulation has strongly influenced its reticence towards rights-based approaches to protecting privacy.[172]

In 1988, the Office of the United Nations High Commissioner for Human Rights announced that Article 17 of the ICCPR was to apply to states, natural persons and legal persons, and that 'all member states are required to adopt legislative and other

163 Joel Reidenberg, 'Resolving Conflicting International Data Privacy Rules in Cyberspace' (2000) 52 Stanford Law Review 1315, 1347.

164 ibid.

165 Schwartz (n 30) 1348.

166 Citron (n 29) 1253.

167 Lindsay (n 40) 160.

168 UDHR, GA Res 217A (III), UN GAOR, UN Doc A/810 (10 December 1948) art 12; ICCPR, opened for signature 16 December 1966, 993 UNTS 3 (entered into force 3 January 1976) art 17.

169 See Human Rights Act 1998 (UK) ch 42; Convention for the Protection of Human Rights and Fundamental Freedoms, opened for signature 4 November 1950, 213 UNTS 221 (entered into force 3 September 1953).

170 Lindsay (n 40) 132.

171 ibid 160.

172 ibid 133.

measures to give effect to the prohibition against such interferences and attacks as well as to the protection of this right'.[173] To satisfy its obligations under the ICCPR, Australia enacted the Privacy Act.[174] However, the Privacy Act is limited to the proper use and access of information and only partially implements Article 17.[175]

In *Australian Broadcasting Commission v Lenah Game Meats (Lenah)*,[176] Justice Callinan advocated for a change to the traditional consequentialist privacy law framework:

> Having regard to current conditions in this country, and developments of law in other common law jurisdictions, the time is ripe for consideration whether ... the legislatures should be left to determine whether provisions for a remedy for [a tort of invasion of privacy] should be made ...[177]

While *Lenah* left the door open for the common law to develop a cause of action for breach of privacy, a judicial majority is yet to take the bold step in developing the action.[178] In 2014, the Australian Law Reform Commission recommended the introduction of a statutory cause of action for privacy intrusion.[179] However, Australia continued to favour a consequentialist approach to information privacy regulation and refused to recognize a right to privacy and a statutory cause of action for privacy invasion. Hence, this article suggests that Australia's traditional consequentialist approach to privacy protection has resulted in a somewhat fragmented approach to privacy protection.[180] In contrast, a deontological approach to information privacy regulation is likely to have supported more effective rights-based proposals to protecting privacy mandated by the ICCPR and UDHR. Davies argues that the main deficiency with information privacy laws based on consequentialist accounts is that they protect data before people.[181] Thus, in order to introduce a right to explanation, Australia's current theoretical approach to information privacy regulation must support a deontological approach which privileges individual autonomy, control and dignity.

The introduction of the APPs into the Privacy Act in 2014, however, evidenced a movement from such a consequentialist approach towards a deontological approach to information privacy regulation.[182] The APPs replaced the former Information

173 Office of the United Nations High Commissioner for Human Rights, *General Comment Number 16: The Right to Respect of Privacy, Family, Home and Correspondence, and Protection of Honour and Reputation*, 32nd sess, UN Doc HRI/GEN/1/Rev.9 (8 April 1988) para 1.

174 Privacy Act (n 104).

175 Lindsay (n 40) 144.

176 (2001) 208 CLR 199.

177 ibid 328 [335] (Callinan J).

178 *Grosse v Purvis* [2003] QDC 151 (Skoien J).

179 Australian Law Reform Commission, *Serious Invasions of Privacy in the Digital Era* (Final Report No 123, September 2014) 9.

180 Lindsay (n 40) 132.

181 Simon Davies, 'Re-Engineering the Right to Privacy: How Privacy Has Been Transformed from a Right to a Commodity' in Philip Agre and Marc Rotenberg (eds), *Technology and Privacy: The New Landscape* (MIT Press, 1997) 143, 156.

182 Niloufer Selvadurai, 'Protecting Online Informational Privacy in a Converged Digital Environment – The Merits of the New Australian Privacy Principles' (2013) 22 Information and Communications Technology Law 299, 306; Privacy Amendment (Enhancing Privacy Protection) Act 2012 (Cth).

Privacy Principles (IPPs) that applied to the handling of personal information by government agencies, and the National Privacy Principles (NPP) that applied to the handling of personal information by large businesses, all health service providers and some small businesses and non-government organizations.[183] The shift towards a deontological approach to information privacy regulation is demonstrated in the increased specificity in how the APPs protect individual identity and autonomy in the digital environment,[184] and the introduction of new prescriptive independent rights such as the right to use a pseudonym[185] and the right to notification.[186]

On 1 August 2019, the Federal Government passed the Treasury Laws Amendment (Consumer Data Right) Bill 2019 (Cth)[187] to introduce the Consumer Data Right (CDR). The CDR establishes an 'economy-wide consumer-directed data transfer system'[188] as it enables individuals to access data that a business holds in relation to them, and to direct the data to be transferred in digital form to third parties and other businesses.[189] The four major Australian banks (ANZ, Westpac, Commonwealth Bank and NAB) started sharing product reference data from July 2019 on a voluntary basis, and consumer data relating to credit and debit cards, deposit accounts and transaction accounts will be made available to individuals from 1 July 2020.

The CDR is directly grounded in a deontological approach to information privacy regulation as it provides individuals greater autonomy and control over how their personal information and data are used by businesses. The 'Explanatory Memorandum' of the Bill states that the CDR 'is designed to give customers more control over their information, leading to more choice in where they take their business, or more convenience in managing their money and services'.[190] The CDR enables consumers to access a broader range of information than the current APP 12.[191] While APP 12 allows individuals to access 'personal information' about themselves,[192] the CDR applies to data that relate to businesses as well as individuals and provides access to information about a service provider's products as well.[193] Hence, the CDR represents an important movement towards a deontological approach to information privacy regulation as it seeks to provide individuals greater data portability rights.[194]

In linking the notion of privacy to autonomy, Westin argued that autonomy means self-government.[195] According to Westin, individuals are self-governing if they

183 Normann Witzleb, 'Halfway or Half-Hearted? An Overview of the Privacy Amendment (Enhancing Privacy Protection) Act 2012 (Cth)' (2013) 41 Australian Business Law Review 55, 55.
184 Selvadurai (n 182) 306.
185 Privacy Act (n 104) sch 1, s 2.
186 ibid, sch 1 s 5.
187 Treasury Laws Amendment (Consumer Data Right) Bill 2019 (Cth).
188 Explanatory Memorandum, Treasury Laws Amendment (Consumer Data Right) Bill 2019 (Cth) 1.
189 ibid 3.
190 ibid.
191 ibid 3.
192 Privacy Act (n 104) sch 1 s 12.
193 Explanatory Memorandum, Treasury Laws Amendment (Consumer Data Right) Bill 2019 (Cth) 3.
194 ibid 1.
195 See Alan Westin, *Privacy and Freedom* (Atheneum 1967).

live their lives freely without being subject to manipulative or invasive external forces.[196] The introduction of the APPs and CDR minimize the opportunities for external interference and manipulation in the digital environment and seek to promote an environment where individuals can freely pursue their lives as they fit.[197] These developments in Australian privacy law are in direct alignment with the European deontological governance philosophy as they demonstrate the government's support for an environment that is conducive to the creation of individual autonomy and control in the digital environment.[198] Therefore, this article submits that a right to explanation should be enacted as it is supported and aligned with Australia's preference for a deontological approach to information privacy regulation.

There is however an important limitation to the value of the GDPR as a model for Australian law reform. The GDPR right to explanation is limited to decisions that are made solely using automated processes. It is suggested that this is overly narrow as even where there is an element of human intervention there is potential for such decisions to be opaque. Moreover, where an individual wishes to challenge such a decision it will not be readily apparent to them what aspect of the decision has been made using automated processes and which component has been made by a human or being the subject of human oversight. As discussed above, a variety of Australian statutes do provide for automated processing in conjunction with human oversight. In such a context, it would be preferable if the decisions to which the right to explanation attaches to were more precisely articulated. In this regard, the description of AI systems in the OECD Recommendations is of assistance. Using this definition as a guide, an 'AI decision' to which a right to explanation attaches could be defined as decisions made using an AI system, defined as 'a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments which are designed to operate with varying levels of autonomy'. Such an expansive definition can enable the right of explanation to provide meaningful assistance to individuals affected by decisions automated by government and commercial entities. Furthermore, while the GDPR does not prescribe the precise form in which explanations should be given, the Australian provision should articulate the matters to be included and the process to be followed. A lack of such guidance could undermine any right to explanation and enable entities to provide broad and generalized information that does not explain the specific basis for the decision in particular circumstances of the individual who is requesting the explanation for the automated decision.

CONCLUSION

As government and commercial entities increasingly rely on automated decisions for their societal efficiency gains, it is necessary to enact a right to explanation and balance the protection of legitimate rights and interests. The absence of such a right can significantly undermine an individual's right to privacy, control and dignity. This article identifies that a right to explanation is inherently grounded in deontological

196 ibid.
197 ibid.
198 Selvadurai (n 182) 306.

approaches to information privacy protection. In this vein, the enactment of a right to explanation in Australia is directly aligned with Australia's movement towards deontological rights-based forms of information privacy protection.

At present, Australian privacy and administrative laws only permit computer-assisted decision-making and do not provide individuals with a right to explanation for such automated or semi-automated decisions. The limitations in such laws further support the enactment of a right to explanation and necessitate a discussion into how to frame such a right. While the European Union GDPR provides a valuable foundation for an analogous right to explanation in Australia, the application of the GDPR right to explanation is narrowly limited to decisions that are wholly automated without any form of human intervention. This article therefore proposes that Australia should adopt the expansive definition of 'AI decision' provided by the OECD Recommendations to enable the right of explanation to provide meaningful assistance to individuals affected by both automated and semi-automated decisions generated by government and commercial entities.

Total transparency of an automated decision-making system is not required to meet the legitimate interests and rights of individuals, and guidance should be provided to government and commercial entities as to the extent of information that must be provided to individuals seeking an explanation, whether that is in the form of counterfactual explanations or broad and generalized information. In this way, Australia should enact a right to explanation for automated decision-making that seeks to calibrate the societal efficiency gains of automated decision-making with the protection of legitimate personal rights and interests.