Castan Centre for Human Rights Law

Faculty of Law, Monash University

Submission to the Human Rights Commission

Discussion Paper on Human Rights and Technology

Prepared by

Dr Maria O'Sullivan, Deputy Director, Castan Centre for Human Rights Law

Dr Yee-Fui Ng, Associate, Castan Centre for Human Rights Law

Dr Normann Witzleb, Associate, Castan Centre for Human Rights Law

Karin Frode, Policy Manager, Castan Centre for Human Rights Law

Andrea Olivares Jones, Project Officer, Castan Centre for Human Rights Law

Estelle Wallingford, PhD Candidate in Artificial Intelligence and the Law

Faculty of Law, Monash University

The Castan Centre for Human Rights Law welcomes the opportunity to make a submission in response to the Human Rights Commission Discussion Paper on Human Rights and Technology. The Castan Centre's mission includes the promotion and protection of human rights. It is from this perspective that we make this submission.

We note that we have targeted our submission on the Discussion Paper to the questions we are best able to answer given the relative areas of expertise by the drafting team. Therefore, not all proposals and questions raised in the Discussion Paper are addressed in this submission.

Part B: Artificial intelligence

Question A: The Commission's proposed definition of 'AI-informed decision making' has the following two elements: there must be a decision that has a legal, or similarly significant, effect for an individual; and AI must have materially assisted in the process of making the decision.

Is the Commission's definition of 'AI-informed decision making' appropriate for the purposes of regulation to protect human rights and other key goals?

(a) Context

The AHRC's definition of AI-informed decision making borrows terms from the European Union's (EU) General Data Protection Regulation (GDPR), specifically Article 22(1) which states:

> *The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.*

Importantly, the prohibition in Article 22(1) of the GDPR provides 'additional safeguards' specific to circumstances of '*solely* automated decision-making' including profiling. The EU's Article 29 Data Protection Working Party Guidelines on Automated Individual Decision-making and Profiling ('EU Guidelines') define such decisions as those made 'by technological means without human involvement',[1] distinguishing them from instances where a human is *meaningfully* involved in decision-making.

[1] European Union, Article 29 Data Protection Working Party, *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679* (2018) 17/EN WP251rev.01, <https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053> ('*EU Guidelines'*), p 8.

EU member states, when supplementing the GDPR in their domestic law, have similarly limited the protections in Article 22(1) to those 'only', 'exclusively', or 'totally' made by automated processing,[2] with the UK emphasising that this is decision-making that 'excludes any human influence on the outcome'.[3] Paul Voigt and Axel Von dem Bussche explain that the application of Article 22(1) therefore only opens up when '*no human* has any decision-making power',[4] irrespective of whether they are otherwise involved in the decision-making process.

The AHRC Discussion Paper seeks to adopt the concept of 'legal and similarly significant effect', which originates from Article 22 to a broader context than the GDPR, where it is limited to decisions 'based solely on automated processing'. Instead, the AHRC proposes to apply the standard to decisions where AI has 'materially assisted in the process of making the decision'. From a human rights standpoint this can be considered both positive and negative.

As for the positive, by extending the scope of protections beyond decisions made 'solely' by AI systems to include decisions where humans are involved, the AHRC envisages regulation that applies in a broader range of circumstances. It would encompass instances such as recruitment decisions where a human is *informed* by an AI system but makes the ultimate, whereas they would not have been under a narrow scope such as that in Article 22(1).

> **Example**: An employer chooses a candidate for employment from a pool of applicants that was created using an algorithm that favoured some people over others. This does not appear to be a decision based 'solely on automated processing' covered under the GDPR but a decision where a human was informed by AI.

A subsequent benefit of wider protections therefore is that they would limit the ability of companies and government bodies to circumvent protections in AI regulation by claiming a human was somehow involved in the process, even if this involvement was trivial or menial.

Unfortunately, however, the threshold created by Article 22(1) of the GDPR through the terms 'legal' and 'similarly significant effects' is relatively high, as it was only intended to apply in very particular circumstances. The EU Guidelines which explore these terms, for example, do not in general consider online advertising to have a legal or similarly significant effect on individuals.[5]

Notwithstanding, the UN Human Rights Council noted in a 2017 Resolution that 'automatic processing of personal data for individual profiling may lead to discrimination or decisions that otherwise have the potential to affect the enjoyment of human rights, including economic, social and cultural rights'.[6] This may include, for example, online targeted advertising which

---

[2] Gianclaudio Malgieri, 'Automated Decision-making in the EU Member States: The Right to Explanation and other "Suitable Safeguards" in the National Legislations' (2019) 35 *Computer and Security Review* 1, 18-19.
[3] Information Commissioner's Office (UK), 'What does the GDPR say about automated decision-making and profiling?', *Information Commissioner's Office* (webpage) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/automated-decision-making-and-profiling/what-does-the-gdpr-say-about-automated-decision-making-and-profiling/> ('*ICO Guidelines*').
[4] Paul Voigt and Axel Von dem Bussche, *The EU General Data Protection Regulation (GDPR): A Practical Guide* (Springer, 2018), p 181 (emphasis in original).
[5] European Union Article 29 Data Protection Working Party, *Guidelines on Automated Individual Decision-making and Profiling for the purposes of Regulation 2016/679*, 17/EN WP251rev.01 (2018) 22 <https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053> ('*EU Guidelines*').
[6] UN Human Rights Council, *The right to privacy in the digital age*, 7 April 2017, A/HRC/RES/34/7 <https://documents-dds-ny.un.org/doc/UNDOC/GEN/G17/086/31/PDF/G1708631.pdf?OpenElement>.

may have significant potential to discriminate, or reflect past prejudice or implicit bias against' protected groups in myriad ways, many of which are not likely to meet the threshold of 'legal' or 'similarly significant'.

> **Example**: Latanya Sweeney of Harvard University in 2013 conducted a study where she investigated search engine advertisements for internet users with names typically associated with African Americans. She found a statistically significant difference in the type of advertisements of African American and non-African American users, with the former far more likely to see advertisements relating to arrest and criminal records than users with Caucasian sounding names.[7]

This raises questions about the appropriateness of borrowing the terms 'legal effects' and 'similarly significant effects':

- Are these terms appropriate in a context which is intended to go beyond decisions based solely on automated processing?; and
- Are the terms too narrow to protect the human rights of the subjects of such a broader range of decisions?

(b) 'Legal' effects

The EU Guidelines provide that a decision with 'legal' effects refers to a decision that affects someone's 'legal rights', 'legal status' or 'rights under a contract'. Further, the Guidelines state that 'only serious impactful effects will be covered by Article 22'.

> **Examples** of legal rights possibly affected in this context include the rights under the Charter of Fundamental Rights of the European Union, such as the right to associate with others, the right to vote, and the freedom to take legal action.
>
> **Examples** of decisions that impact upon legal status or contractual agreements in turn include decision impacting on admission into a country, denying citizenship, affecting entitlements to social benefits granted by law, and cancellation of contracts.

The UK Information Commissioner has similarly noted that the 'legal' effects referred to in the GDPR must 'adversely affec[t] someone's legal rights'.[8] Voigt and Von dem Bussche emphasise in their 'Practical Guide to the GDPR' that this includes both positive and negative effects for the data subject.[9]

---

[7] Latanya Sweeney, 'Discrimination and Online Ad Delivery' (2013) 50(5) *ACM Queue,* 44-54; See Lillian Edwards and Michael Veale, 'Slave to the Algorithm? Why a Right to an Explanation' is Probably not the Remedy You are Looking For' (2017) *Duke Law & Technology Review* 16(1) 46; See also Katarina Throssell, 'When Algorithms Discriminate: A Framework for the Ethical Use of Algorithmic Decision-Making in the Public Sector', *Department of Premier and Cabinet* (Report, 23 May 2018) 22.

[8] Information Commissioner's Office, 'Rights related to automated decision making including profiling', *Information Commissioner's Office* (webpage) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/rights-related-to-automated-decision-making-including-profiling/>.

[9] Paul Voigt and Axel von dem Bussche, T*he EU General Data Protection Regulation (GDPR): A Practical Guide* (Springer Press, 2018) 182.

The benefit of interpreting 'legal' effects in this way is that 'impacts on legal status can be determined according to the letter of the law', whereas decisions that are 'significant' are much vague and open to perception, as discussed below.[10]

It should however be noted that 'legal rights' in the context of the European Union differs to that in Australia. In particular, the existence of a European Charter of Human Rights establishes binding fundamental rights for EU citizens. With such human rights enshrined in law, automated decisions that impact upon these freedoms could be considered to have a 'legal effect'.

Conversely, the absence of such protections in most Australian jurisdictions through a similar human rights instrument limits the utility of protections for subjects impacted by AI decisions with 'legal effect'. The introduction of a federal charter of rights would therefore be an important step to enhancing the protections provided by AI regulation.

(c) 'Similarly significant effects'

There has been considerable contention around the use of the phrase 'similarly significant effects'.

Firstly, academics have raised concerns around the vagueness of the term 'significant' in the context of Article 22 of the GDPR. Elena Gil Gonzales Sandra Wachter, Brent Mittelstadt and Luciano Floridi, for example, question what perspective should be taken into consideration when defining significant effects - should such effects be significant from the subjective perspective of the data subject? Or measured by an external standard?[11]

The EU Guidelines provide some guidance on the use of the term within the context of the GDPR. Firstly, they state that 'significant' means 'sufficiently great or important to be worthy of attention'. This does little to elucidate the meaning of the term as it does not answer the question of what is 'sufficient', or 'important', nor the question of worthy of attention to *whom?*

Commentary from the Guidelines appears to indicate an objective standard. The EU specifies that such decisions must:

i. Impact upon the circumstance, behaviour or choices of the data subjects;
ii. Have a prolonged or permanent impact on the data subject; or
iii. Lead to the exclusion or discrimination of individuals.

Other commentary in the Guidelines, however, suggests that 'significance' can be subjective in certain circumstances, stating that 'processing that might have little impact on individuals generally may in fact have a significant effect for certain groups of society, such as minority groups or vulnerable adults.'[12] Further, the Guidelines state that children require enhanced protection.

---

[10] Sandra Wachter, Brent Mittelstadt and Luciano Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) *International Data Privacy Law* 7(2) 76, 92-93; See also Lee Bygrave, 'Minding the Machine: Art 15 of the EC Data Protection Directive and Automated Profiling' (2000) *Privacy Law and Policy Reporter* 7(4) 67.
[11] Wachter, Mittelstadt and Floridi, above n 10, 98.
[12] EU Guidelines p. 22; See also Emily Pehrsson, 'The Meaning of the GDPR Article 22', (2018) 31 *European Law Working Papers* 1, 16 <https://law.stanford.edu/wp-content/uploads/2018/05/pehrsson_eulawwp31.pdf.

As raised by Emily Pehrsson of Stanford University, this continued ambiguity 'could immerse the courts in a slew of litigation and hurt business across the EU by creating unpredictability in Article 22's application'.[13]

Examples of decisions that have similarly significant effects provided by the Guidelines include decisions that affect a subject's financial circumstances, health services, employment opportunity or access to education. EU Recital 71 specifically names instances of online credit eligibility assessments and online employment recruiting as decisions which have 'similarly significant effects'.[14]

Academic commentary indicates that 'similar effects' are those which produce negative personal or economic consequences for the subject.[15] These, Voigt and Von dem Bussche explain, must be determined on a 'case-by-case basis'.[16]

It should be noted that the introduction of the word 'similarly' is new to the GDPR - the word was not present in the GDPR's predecessor the European Data Protection Directive. The EU Guidelines state that it's inclusion indicates that the 'threshold for significance must be similar to that of a decision producing a legal effect'. Therefore, 'similarly significant effects' are understood in the GDPR to mean those in which there is:

(i) no change to legal rights or obligations; and
(ii) the data subject is *still* impacted sufficiently so as to require protection.

The inclusion of the word 'similarly' may not be advisable in the Australian context. Firstly, introducing a threshold of significance elevated to a standard akin to a legal right may create barriers to justice for vulnerable individuals not currently protected under existing legislation. As mentioned above for example, the absence of a federal charter of rights limits the grounds upon which subjects impacted by AI-informed decision-making make claim a decision has had 'legal effect'.

Further, AI regulation in Australia may benefit from the creation of similar guidelines to those in the EU, which further expand upon examples of AI-informed decision-making that would be considered to have a 'significant' effect.

(d) For an individual

Another potential issue is the sole inclusion of 'individuals' as the subjects of AI-informed decisions in the AHRC's proposed definition. Edwards and Veale for example raise that this focus on individuals a may result in a lack of protection for groups that may be as affected by a decision as an individual might.[17] They cite the following example:

> **Example:** In 2004, Google search engine algorithms, which ranked results for search queries, placed the site "Jew Watch" as the top ranking for searches for the word "Jew". This site was in fact an anti-Semitic website and had been so highly ranked by the algorithm because the word "Jew" was found too often be used in an anti-Semitic context.

[13] Pehrsson, *The Meaning of the GDPR Article 22,* above n 12.
[14] EU GDPR Recital 71, <https://www.privacy-regulation.eu/en/r71.htm>.
[15] Voigt and Von dem Bussche, above n 4, 182.
[16] Ibid.
[17] Edwards and Veale, above n 7, 48.

Edwards and Veale contend that whilst this may have impacted upon some individuals, it more likely impacted upon a larger group. Their contention is supported by the EU Guidelines, which note that 'decisions that have little impact on individuals generally may have a significant effect for certain groups of society, such as minority groups and vulnerable adults'.

A parallel may perhaps be drawn to the definitions of direct and indirect discrimination where the AHRC recognises, for example, that direct discrimination happens when "a person, or a group of people, is treated less favourably than another person or group because of their background or certain personal characteristics".[18] Similarly, the AHRC recognises that "[i]t is also discrimination when an unreasonable rule or policy applies to everyone but has the effect of disadvantaging some people because of a personal characteristic they share".[19]

Therefore, the inclusion of 'groups' as the subject of AI-informed decision-making would be advisable from a human rights perspective.

(e) Decision-making

The use of the term 'decision making', as opposed to 'decisions' appears to be in line with academic commentary on the subject.

This is primarily for two reasons, the first being that there is some contention as to whether AI systems *can* produce legitimate 'decisions'. Gloria Phillips-Wren for example contends that AI can only 'attemp[t] to mimic human decision-making in some capacity'.[20] Lillian Edwards and Michael Veale similarly raise that while AI systems can produce outputs as classifications or estimations, they are still 'incapable of synthesising the estimation and relevant uncertainties into a *decision* for action'.[21]

Beyond academia, in Australia, courts have adopted a similar view. In the case of *Pintarich,* in which AI enabled automated systems were used to calculate and claim social services debt for welfare recipients, the Federal Court of Australia found that no 'decision' was made 'unless accompanied by the requisite mental process of an authorised officer'.[22] This means that at least for administrative decisions, a human must be involved for a computed decision to constitute a legal decision.

Secondly, it is also prudent, in the development of regulation, to also generate protections for decisions not solely made by AI, but meaningfully assisted or impacted by AI.

---

[18] Australian Human Rights Commission, 'Discrimination', *Australian Human Rights Commission* (webpage) <https://www.humanrights.gov.au/quick-guide/12030>.
[19] Ibid.
[20] Gloria Phillips Wren, 'Intelligent Decision Support Systems', in Gloria Phillips-Wren, Nikhil Ichalkaranje and Lakhmi Jain, *Intelligent Decision Making, an AI-Based Approach* (Springer Press, 2008) 1.
[21] Edwards and Veale, above n 7, 46.
[22] *Pintarich v Deputy Commissioner of Taxation* [2018] FCAFC 79; discussed in Yee-Fui Ng and Maria O'Sullivan, 'Deliberation and Automation – When is a Decision a 'Decision'?' (2019) 26(1) *Australian Journal of Administrative Law* 21-34; see also Kobi Leins, 'What is the Law When AI Makes Decisions?', *University of Melbourne Pursuit* (Blog) <https://pursuit.unimelb.edu.au/articles/what-is-the-law-when-ai-makes-the-decisions>.

Recommendations

- In relation to ‘legal effects’: the introduction of a federal charter of rights and comprehensive equality law will be an important step in enhancing the protections provided by AI regulation.

- ‘Similarly significant effect’: we recommend that the word ‘similarly’ be removed as it adds another layer of confusion. There is a risk that ‘similarly’ could be used to limit rights protection and may lead to lengthy arguments in court about whether or not an effect is similarly significant to a legal effect.

- Any regulation must be accompanied by clear and accessible guidelines for duty-bearers and rights-holders with concrete examples .

Proposal 4: The Australian Government should introduce a statutory cause of action for serious invasion of privacy.

The Castan Centre for Human Rights strongly support this proposal.

I. Introduction

Over the years, numerous Australian law reform bodies recommended the introduction of a statutory cause of action for invasion of privacy. Yet, despite the unanimous support for legislative action by successive law reform inquiries, victims of privacy invasion in Australia still need to rely on a patchwork of general law and statutory provisions that protect aspects of privacy incidentally, rather than through a single, comprehensive right to privacy. In July 2019, the Australian Competition and Consumer Commission (ACCC) added its voice in support of the introduction of a statutory cause of action for serious invasions of privacy.[23] In the Final Report of its major inquiry into *Digital Platforms*, the ACCC proposes that the statutory privacy tort should be enacted in the form that had been recommended by the Australian Law Reform Commission already in 2014.[24] In December 2019, in the context of this inquiry, the Australian Human Rights Commission (AHRC) also urged that this ALRC recommendation be implemented.

The ACCC Digital Platforms inquiry also support a statutory cause of action

Although both Commissions made their proposals in different contexts, their respective calls for legislative action demonstrate the common threat that the rise of modern data-driven technology poses for individual privacy. The AHRC expressed the expectation that a statutory privacy tort could address concerns about the potential misuse of personal information in the context of decision making informed by artificial intelligence.[25] The ACCC Digital Platforms Inquiry examined the adequacy of Australian regulation of digital platforms in light of their transforming impact on the news media and advertising sector. Data protection and privacy laws were just one aspect of a broad-ranging inquiry that also included competition law, media law and consumer protection law. Given this focus, the ACCC was persuaded that a statutory cause of action would increase the accountability of businesses for their data practices and give consumers greater control over their personal information. Related recommendations in the report were directed at strengthening the *Privacy Act 1988* (Cth), including by broadening its scope, enhancing consent requirements, increased penalties for contraventions, and by introducing a direct right of action for those who suffer an interference with their information privacy rights under the Privacy Act.

[23] See Australian Competition and Consumer Commission, *Digital Platforms Inquiry*, Final Report, July 2019, Recommendation 19.
[24] Australian Law Reform Commission, *Serious Invasions of Privacy in the Digital Era*, Report 123, 2014.
[25] Australian Human Rights Commission, *Human Rights and Technology*, Discussion Paper, 92.

The Government has responded to the ACCC reform proposals in December 2019 by reiterating its earlier commitment to amend the Privacy Act to increase penalties, strengthen enforcement and requiring social media platforms to subscribe to a binding privacy code.[26] In relation to the call for a statutory cause of action, the Government announced that this recommendation would be examined as part of a 'review of the Privacy Act and related laws to consider whether broader reform of the Australian privacy law framework is necessary in the medium- to long-term to empower consumers, protect their data and best serve the Australian economy'.[27]

The renewed attention given to privacy law reform at federal level, especially to a statutory cause of action, makes it timely to contextualise and evaluate the recent ACCC and AHRC proposals for a statutory privacy tort. The next part of the submission will explain the design of the cause of action proposed by the ALRC, as recommended for adoption by the ACCC and the AHRC. Part III will make the case for a statutory tort by explaining why legislative action is preferable over awaiting further developments of privacy protections by the courts. Part IV will identify some shortcomings of the ALRC proposal that should still be addressed in the further law reform process. Part V will provide a summary of our support for proposal 4.

II. The ALRC statutory privacy tort

In line with its terms of reference, the ALRC inquiry into *Serious Invasions of Privacy in the Digital Era* did not engage in the debate on whether statutory reform is preferable over judicial development of privacy protection. That question had already been answered in the affirmative by the ALRC's broad enquiry into privacy law reform in 2007-2008.[28] Instead, the 2013 enquiry ALRC was tasked with considering *how* a statutory cause of action would best be formulated. In its inquiry, the Commission engaged with similar proposals for enhanced civil redress for privacy breaches made by the New South Wales Law Reform Commission in 2009[29] and the Victorian Law Reform Commission in 2010.[30] While these other inquiries traversed much the same ground and the recommendations put forward share a large number of similarities, there are also some important differences.[31]

---

[26] Australian Government, *Regulating in the Digital Age: Government Response and Implementation Roadmap for the Digital Platforms Inquiry*.
[27] Australian Government, *Regulating in the digital age: Government Response and Implementation Roadmap for the Digital Platforms Inquiry* (2019), p. 18.
[28] Australian Law Reform Commission, For Your Information: Australian Privacy Law and Practice (Report 108), 2008.
[29] New South Wales Law Reform Commission, *Invasion of Privacy*, Report 120, 2009.
[30] Victorian Law Reform Commission, Surveillance in Public Places, Final Report 18, 2010). See also Law Reform Committee of the Victorian Parliament, *Report of Inquiry into Sexting* (May 2013), which endorsed the recommendation of the VLRC.
[31] For more detailed comparison, see Barbara McDonald, 'A statutory action for breach of privacy: Would it make a (beneficial) difference?' (2013) 36 *Australian Bar Review* 241 (Professor McDonald was the Commissioner in charge of the ALRC privacy reference); Des Butler, 'Protecting personal privacy in Australia: Quo vadis?' (2016) 42 *Australian Bar Review* 107; David Lindsay, 'A privacy tort for Australia? A critical appreciation of the ALRC report on serious invasions of privacy' (2015) 12 *Privacy Law Bulletin* 8. For a comparison of the proposal prior to the 2013 ALRC report: Normann Witzleb, 'A statutory cause of action for privacy? A critical appraisal of three recent Australian law reform proposals' (2011) 19 *Torts Law Journal* 104

Before formulating its preferred design, the ALRC considered the available options in detail and consulted widely with stakeholders and the community, who overwhelmingly supported legislation. However, when the ALRC presented its report, the government of the day, led by Tony Abbott, made clear that it did not accept the recommendations made.[32] The rejection of statutory law reform was in keeping with the stance of successive Australian federal governments not to legislate for a general right to privacy. In response to the lack of positive action at federal level, a number of state-based law reform inquiries subsequently recommended legislation at state level. The NSW Legislative Council Standing Committee on Law and Justice proposed in 2016 that NSW introduce a statutory privacy tort that should be based largely on the ALRC model.[33] In the same year, a relatively little known inquiry by the South Australian Law Reform Institute came to a similar proposal for a state-based approach and recommended the establishment of a South Australian civil law action for serious invasion of personal privacy.[34]

The design of the ALRC tort

The ALRC has proposed federal legislation creating a new tort of serious invasion of privacy with the following characteristics:

- The tort is limited to 'intrusion into seclusion' and 'misuse of private information'. 'Intrusion' includes activities such as physically intruding into the plaintiff's private space or by watching, listening to or recording the plaintiff's private activities or private affairs.[35] A 'misuse' occurs by activities such as collecting or disclosing private information about the plaintiff.[36] By confining the torts to these two major scenarios of privacy interference, the ALRC steered a middle course between a broad and potentially open-ended privacy action[37] and proposals that sought to create two separate torts for intrusion and misuse.[38]
- The new tort is actionable only where a person in the position of the plaintiff had a reasonable expectation of privacy, in all of the circumstances. This requirement adopts an internationally accepted threshold test of when a person's right to privacy is engaged.[39]
- The tort is confined to intentional or reckless invasions of privacy, so that merely negligent invasions of privacy would not become actionable. In doing so, the ALRC formulated the cause of action more narrowly than prior proposals by the NSWLRC and the VLRC. This limitation has subsequently been criticised and will be considered in more detail below in Part 4.

[32] Chris Merritt, 'Brandis rejects privacy tort call', *The Australian*, 3 April 2014.
[33] NSW Legislative Council Standing Committee on Law and Justice, *Remedies for the serious invasion of privacy in New South Wales* (Report no. 57, 2016).
[34] South Australian Law Reform Institute, *Too much information: A statutory cause of action for invasion of privacy* (Final Report 4), 2016.
[35] Australian Law Reform Commission (above n 2), Rec 5-1 (a).
[36] Ibid, Rec 5-1 (b).
[37] This was proposed by the New South Wales Law Reform Commission (above n 7), 4.14.
[38] This was proposed by the Victorian Law Reform Commission (above n 8), Rec. 22.
[39] See Australian Government, *A Commonwealth Statutory Cause of Action for Serious Invasion of Privacy*, Issues Paper (2011), pp. 17-21.

- The scope of the tort is further limited by introducing a threshold requirement that the invasion must be serious. 'Seriousness' can be established by reference to the degree of any offence, distress or harm to dignity caused, or the motivation of the defendant, in particular malice, or by reference to other relevant factors.[40] This requirement did not originate from the findings of the ALRC, but was part of the terms of reference, which tasked the ALRC to enquire specifically into the remedies for *serious* invasions of privacy.
- Lastly, the ALRC proposed that an action could only succeed if the court was satisfied that the public interest in privacy outweighs any countervailing public interests. This requirement for a balancing exercise seeks to ensure that conflicting interests such as freedom of speech, freedom of the media, public health and safety, and national security are not disproportionately curtailed. While the defendant has an evidential burden in relation to these matters, it is part of the plaintiff's case to make out that the interest in privacy outweighs countervailing public interests.[41]

If a plaintiff establishes that the tort – as defined above – has occurred, a defendant can rely on a number of defences and exemptions, such as consent, necessity, absolute privilege, and fair report of proceedings of public concern.[42] The report recommends that a broad range of remedies should be available to successful privacy claimants.[43] These include traditional tort remedies such as damages, including compensation for emotional distress, injunctions and an account of profits. In exceptional circumstances, a court would be empowered to award exemplary damages; however, the ALRC envisaged a cap on the total amount of damages.[44] In addition, the report also recommends to give the court the power to make orders that are more specifically directed at remedying privacy harms, such as declarations, orders for apologies and corrections.

Only some further amendments are needed

In its inquiry, the ALRC carefully evaluated the existing law and engaged in extensive community consultation. The model tort for a statutory tort it arrived at seeks to balance the various interests that on collide in cases of privacy invasions. Although not implemented, the ALRC model has become the reference point for all subsequent debates of how a civil right of redress for privacy invasion should be formulated. Subject to the further amendments suggested below at IV, we submit that the ALRC tort for serious invasion of privacy should be adopted.

III. Why a statutory tort?

It has now been nearly 20 years since the High Court declared in *Australian Broadcasting Corporation v Lenah Game Meats Pty Ltd*[45] that there are no obstacles to the recognition of a

[40] Australian Law Reform Commission (above n 24), Rec 8-1.
[41] Australian Law Reform Commission (above n 24), [9.77].
[42] Ibid, ch 11.
[43] Ibid, ch 12.
[44] Ibid, rec 12-5.
[45] *Australian Broadcasting Corporation v Lenah Game Meats Pty Ltd* (2001) 208 CLR 199; [2001] HCA 63.

common law right to privacy. Yet, despite this assurance, no Australian appellate court has to date has seen fit to recognise the existence of a privacy tort. In the courts, the law of privacy protection appears to not have moved significantly beyond the 2008 decision of the Victorian Court of Appeal in *Giller v Procopets*.[46] In that case, the plaintiff was the victim of serious intimate image abuse following the breakdown of a long-term relationship with the defendant. However, the Court considered it unnecessary to decide whether such a generalised tort of invasion of privacy should be recognised.[47] It was content to protect the plaintiff's interests on the basis of a claim for breach of confidence and, in doing so, recognised for the first time that equitable compensation following a breach of personal confidence can include an award to compensate for non-pecuniary harm, in particular injury to feelings.[48]

Other causes of action leave gaps

Despite this extension of the remedial options, breach of confidence is only partially suited to the task of responding to privacy invasions. The well-known limitations of this cause of action include that, at least in its original form, it is concerned with protecting relationships of confidentiality, rather than private information per se. While courts have been prepared to extend the scope of the equitable cause of action to cases in which a defendant surreptitiously obtained private information, it remains an open question how broadly it can operate in circumstances in which no prior relationship between the parties exists. Even more importantly, while breach of confidence can deal with many instances of unauthorised disclosure of personal information, it is not designed to protect against the mere intrusion into the personal sphere (that is not accompanied by the misuse of personal information). The besetting, surveillance and stalking of persons may in some cases lead to liability under other torts, such as trespass to land or nuisance, or constitute a criminal offence under surveillance legislation or so-called 'upskirting' laws, but the protection offered by these mechanisms is piecemeal, often not primarily directed at the protection of privacy, and leaves some gaps.[49]

The courts are unlikely to recognise a common law right to privacy

A statutory privacy tort would help overcome to reluctance of Australian courts to recognise a right to privacy and would ensure that Australia's privacy protection no longer lags behind its counterparts in other common law jurisdictions. Australia is now virtually unique among major common law jurisdictions in not recognising a legally enforceable right to privacy. In the majority of comparable jurisdictions, privacy protections have been developed through the courts. This has been the case in the United Kingdom, New Zealand and, more recently, in Canada. In all these countries, a bill of rights or other human rights legislation has provided a framework for the judicial development of a cause of action to protect privacy. Often, courts have been prompted to recognise a common law right to privacy by considering human rights legislation which guarantees a right to respect for private life alongside other fundamental

[46] *Giller v Procopets* (2008) 40 Fam LR 378; [2008] VSCA 236.
[47] Ibid, at [167]-[168] (Ashley JA) and [447]-[452] (Neave JA, Maxwell P agreeing).
[48] The plaintiff was awarded $50 000 damages (including aggravated damages) for mental distress.
[49] Australian Law Reform Commission (above n 24), ch 3.

freedoms, including the right to freedom of expression. In Australia, however, the absence of a federal human rights instrument has stultified the development of a common law right to privacy. It is that gap in the law that the proposed statutory privacy tort would close.

If we look at developments in other common law jurisdictions, we see that courts elsewhere have taken a much more active role. The USA have for many years accepted the existence of privacy torts. In the classification of the Restatement of the Law of Torts (2nd),[50] which in turn accepted the classification by American torts scholar, Professor Dean Prosser, they are:

1. Intrusion upon the plaintiff's seclusion or solitude into his private affairs.
2. Public disclosure of embarrassing private facts about the plaintiff.
3. Publicity which places the plaintiff in a false light in the public eye.
4. Appropriation, for the defendant's advantage, of the plaintiff's name or likeness.[51]

Equally, the United Kingdom, New Zealand and Canada have left Australia behind in this area and enhanced the protection of privacy at common law. The prime example for the assistance that a human right of privacy can provide for the development of domestic law is the UK. The *Human Rights Act 1998* (UK) was intended to give the provisions of the European Convention on Human Rights domestic effect. Soon after the Act came into force in 2000, the courts responded to the new environment by enhancing privacy protection at general law.[52] Initially, they expanded the traditional action for breach of confidence, and then – after the decision in *Campbell v MGN* by the House of Lords – they expressly acknowledged the existence of a new tort of misuse of private information.[53] This tort is now well-accepted in the UK, but still maintains its close links with the rights afforded under the European Convention on Human Rights, as is particularly evident from the elements of this tort, which establish a two-stage enquiry:

(a) If a claimant can establish reasonable expectation of privacy then the right to respect for private life in art.8 of the ECHR is 'engaged' and the first hurdle in the misuse of private information action is cleared.
(b) In the second stage, it is then up to the defendant to show that that right is outweighed by some other interest, usually the right to freedom of expression guaranteed by art.10 of the ECHR.[54]

The situation in New Zealand and Canada is comparable, although the path to recognition of a privacy tort was somewhat different. In both jurisdictions, the effect of human rights law was more indirect because neither the *Canadian Charter of Rights and Freedoms* nor the *New Zealand Bill of Rights Act 1990* contain a broad right to respect for private life, as under European human rights law or the ICCPR. Instead, these instruments provide more limited protection against 'unreasonable search and seizure'. The human rights framework was nonetheless an important driver of law reform. In Canada, four provinces had already established a statutory privacy tort when, in 2012, the Ontario Court of Appeal recognised in *Jones v Tsige*

[50] American Law Institute, *Restatement of the Law (Second) of Torts* (1977) § 652A.
[51] William L Prosser, 'Privacy' (1960) 48 *California Law Review* 383, 389.
[52] *Douglas v Hello! Ltd* [2001] QB 967.
[53] *Campbell v MGN Ltd* [2004] 2 AC 457.
[54] *McKennitt v Ash* [2008] QB 73.

before the tort of intrusion into seclusion.[55] In a subsequent case in 2016, the Ontario Superior Court of Justice recognised, for the first time in Canada, the privacy tort of 'publication of embarrassing private facts'.[56]

The New Zealand *Bill of Rights Act 1990* also does not explicitly recognise a right to privacy, but merely a right to be secure against unreasonable search and seizure. Nonetheless, the New Zealand Court of Appeal engaged in a detailed analysis of the human rights context, when it recognised, in *Hosking v Runting*,[57] the existence of a cause of action protecting in relation to publicising private information. In 2012, the New Zealand High Court further developed the law when it accepted, for the first time, the existence of a tort against privacy intrusion in the case of *C v Holland*.[58]

Australia lacks a human right to privacy

Australia has the disadvantage that it does not have a constitutional bill of rights at federal level, but only state and territory human rights legislation in Victoria, the ACT and now Queensland. Australia, like most countries, is a signatory of the International Covenant on Civil and Political Rights (ICCPR), which in its Art. 17 imposes on state parties an obligation to protect everyone against arbitrary or unlawful interference with their privacy, family, home or correspondence. However, the ICCPR does not form part of domestic Australian law, and Australian courts are reluctant approach to develop the common law in line with international human rights obligations. More generally, Australian judges also appear to feel discomfort at the prospect of recognising relatively high level concepts as the basis of new rights. This is apparent not only in the context of privacy but, for example, also from the reluctance of embracing concepts such as unjust enrichment[59] or good faith in contract law.[60] These factors, combined with the strong media opposition against any expansion of privacy claim rights, dampen any expectations that Australian courts would recognise and develop a comprehensive right to privacy. The notorious lack of political action would also pose an obstacle for any courts that was sympathetic to recognising the right to privacy. It would be easy to denounce any common law right to privacy as the product of illegitimate judicial activism and as lacking democratic legitimacy, when so many calls for parliamentary action have gone unheeded.

[55] *Jones v Tsige* (2012) 108 OR (3d) 241 (CA). For further discussion, see Jeff Berryman, 'Remedies for Breach of Privacy in Canada', in Jason N E Varuhas and Nicole A Moreham (eds), *Remedies for Breach of Privacy* (Oxford: Hart Publishing, 2018), p. 323.
[56] *Jane Doe 464533 v ND* (2016) 128 OR (3d) 352 (Sup Ct J).
[57] *Hosking v Runting* [2005] 1 NZLR 1.
[58] *C v Holland* [2012] NZHC 2155. For further discussion, see Chris DL Hunt, 'New Zealand's New Privacy Tort in Comparative Perspective', (2013) 13 *Oxford University Commonwealth Law Journal* 157.
[59] *Mann v Paterson Constructions Pty Ltd* [2019] HCA 32, [79] (Gageler J), [119] (Nettle, Gordon, Edelman JJ); *Roxborough v Rothmans of Pall Mall Australia Ltd* (2001) 208 CLR 516 at 543-544 [71]-[73] (Gummow J).
[60] The High Court has so far left open whether a general obligation to act in good faith in the performance of contracts should be recognised: *Commonwealth Bank of Australia v Barker* (2014) 253 CLR 169; see also Jeannie M Paterson, 'Good Faith Duties in Contract Performance' (2014) 14 *Oxford University Commonwealth Law Journal* 283.

The counterarguments hold insufficient weight

One common argument put forward against the enactment of a statutory privacy tort is that, in light of existing protections at general and statute law, there is no demonstrated need for it. However, the weight of submissions to previous enquiries suggests that the majority of stakeholders, and the community broadly, have valid concerns about increasing threats to privacy and would prefer a tort to be enacted. Similarly, the AHRC inquiry identifies the issue that technological process in data driven societies creates major news risks for the right to privacy, such as those arising from new technologies such as AI, facial recognition and big data analytics.

The second argument relates to the concern that a statutory privacy tort has the potential to stifle media expression. However, on closer consideration this argument also lacks force. The ALRC was at pains to limit the scope of the privacy tort and to protect media freedom. It has therefore been persuasively argued that the interests of the media are better protected under the ALRC model than they are under current law,[61] which is uncertain and does not adequately address the potential conflict between privacy rights and media freedoms. Consistently with this, the ACCC has pointed out that countervailing public interest matters were 'carefully considered and addressed by the ALRC in designing its statutory cause of action' for privacy.[62]

There are a number of mechanisms to ensure that that a defendant's legitimate interests are sufficiently protected. First, the tort is narrowly defined because it requires intentional or reckless conduct, and that it introduces a threshold requirement of a *serious* invasions of privacy. Second, the 'seriousness threshold' operates in addition to the public interest balancing test, a construction which the ALRC acknowledges was intended to 'further ensure the new tort does not unduly burden competing interests such as freedom of speech'.[63] It has been argued that this design feature has the potential to cause 'duplication'[64] and may not be necessary to deter or exclude trivial claims. Third, the ALRC purposefully made it part of the plaintiff's case to demonstrate that the public interest in privacy outweighs the public interest in freedom of expression. This means not only that it is the plaintiff who carries the ultimate burden of establishing that the interest in privacy should prevail over other public interests. By requiring that the *public* interest in privacy outweigh other public interests, the plaintiff must also establish that her private interest in maintaining her privacy coincides with a corresponding public interest. This has the potential to exclude or discount any interest in privacy that does not transcend into the public domain. In their interplay, these features of the ALRC tort ensure that the legitimate interests of others are more than adequately protected. Indeed, as we submit below in IV, consideration should be given to proposals to widen the scope of protection somewhat.

In conclusion, we submit that a statutory privacy tort that is custom built to respond to the tensions existing between the right to privacy and other rights and interest is likely to enhance our freedoms rather than curtail them. It cannot be denied that a privacy tort may on occasion limit freedom of speech – to some extent, that is precisely its point. However, the introduction of a statutory privacy tort with finely calibrated defences would ensure that this occurs only where

[61] Paul Wragg, 'Enhancing Press Freedom through Greater Privacy Law: A UK Perspective on an Australian Privacy Tort' (2014) 36 *Sydney Law Review* 619, 622.
[62] Australian Competition and Consumer Commission (above n 23), 494–495.
[63] ALRC, above n 24, [8.15].
[64] Lindsay, above n 31, 10.

the significance of a person's privacy demonstrably outweighs conflicting public interests, including the interest in free speech.

IV. Proposed amendments to the ALRC privacy tort

The ALRC proposal has many strengths, but it also has a few shortcomings that a future law reform process could still seek to address. A major concern with the ALRC privacy tort is that it is proposed to be limited to intentional and reckless invasions of privacy.[65] In contrast, neither the Victorian Law Reform Commission[66] nor the NSW Law Reform Commission[67] recommended in their reports on privacy to establish a fault standard that excluded negligence. While both commissions anticipated that most actionable invasions of privacy would be committed with intention or recklessness they preferred to retain the option that, in exceptional cases, a negligent invasion of privacy could also be actionable.

The NSW Legislative Council Standing Committee on Law and Justice proposed a compromise model. In its 2016 report, the Committee recommended that NSW introduce a statutory privacy tort that should be based largely on the ALRC model,[68] but that consideration be given to 'incorporating a fault element of intent, recklessness and negligence for governments and corporations, and a fault element of intent and recklessness for natural persons'.[69] We support this recommendation and submit that it should be further considered in the reform process.

A fault element of intent and recklessness for natural persons would mean that an individual would not incur liability if, say, he or she unintentionally encroaches into another's private sphere or thoughtlessly posts on social media sites of photographs depicting friends, family or strangers in embarrassing situations. Limiting this carve-out for negligence to individuals would, however, ensure that corporations would be held to a higher standard. Corporate actors would remain liable for conduct that fails to comply with a standard of reasonable care. In that way, media organisations would be required to engage in responsible journalism that has proper regard to legitimate claims for privacy. Government entities would incur liability when they fail to put in place reasonable security safeguards to protect private information against unauthorised access or loss.[70]

A differentiation between individuals and corporations would also respond to the concern acknowledged by the ALRC that there should be adequate deterrence, and remedies, against data breaches by commercial and government entities.[71] With the greater potential of many business

---

[65] Ibid, Recommendation 7-1.
[66] Victorian Law Reform Commission, *Surveillance in Public Places*, Final Report 18 (2009).
[67] NSW Law Reform Commission, *Invasion of Privacy*, Report 120 (2009), [6.9].
[68] NSW Legislative Council Standing Committee on Law and Justice, *Remedies for the serious invasion of privacy in New South Wales* (Report no. 57, 2016), recs 3 and 4.
[69] Ibid, rec 5.
[70] See, for example, Office of the Australian Information Commissioner, *Department of Immigration and Border Protection: Own motion investigation report*, 1 November 2014, https://www.oaic.gov.au/privacy/privacy-decisions/investigation-reports/department-of-immigration-and-border-protection-own-motion-investigation-report/
[71] ALRC, above n 24, [7.66].

and government entities to commit significant privacy breaches, they should also have greater responsibilities to guard against them. In addition, corporations will often have better resources to make sure (e.g. through training of officers and employees or seeking professional advice) that their practices comply with accepted standards and community expectations on privacy safeguards. Corporations will also generally find it easier to carry the burden of liability for breach, such as through public liability insurance, professional indemnity insurance or pricing mechanisms.

V. Conclusion on Proposal 4

The Castan Centre for Human Rights strongly supports Proposal 4 that the Australian Government should introduce a statutory cause of action for serious invasion of privacy.

The possible arguments have long been exchanged, and – as discussed above – review after review has come down in favour of implementing a new privacy tort. The need for a tort has been widely accepted by most stakeholders and would bring Australia into line with the protections available in comparable common law jurisdictions. Since the ALRC made its well-considered and well-received proposal for a statutory privacy tort, even the design of the cause of action has now been quite firmly in place. While some media interests continue to oppose a privacy tort, the concerns that such a tort might unduly inhibit press freedom have been carefully considered and addressed by the ALRC model. The only ingredient still missing is the political will to introduce it.

The upcoming review of the Australian privacy law framework that has already announced by the Australian government would provide a mechanism in which the design of a statutory privacy tort could be examined once more. However, given the pace of technological development and the ever-increasing potential for the misuse of personal information and other privacy-invasive practices, the need for a tort for the protection of privacy can no longer seriously be doubted.

Proposal 7: The Australian Government should introduce legislation regarding the explainability of AI-informed decision making. This legislation should make clear that, if an individual would have been entitled to an explanation of the decision were it not made using AI, the individual should be able to demand:

(a) a non-technical explanation of the AI-informed decision, which would be comprehensible by a lay person, and

(b) a technical explanation of the AI-informed decision that can be assessed and validated by a person with relevant technical expertise."

We support proposal 7 that recommends that the Australian Government introduce legislation regard the explainability of AI-informed decision-making, including both a technical and non-technical explanation of the AI-informed decision.

In particular, we agree with proposal 7 requiring a 'non-technical explanation of the AI-informed decision, which would be comprehensible by a lay person', as this will enable the non-technical layperson to more readily comprehend the decision affecting them. The requirement for the Australian Government to provide an explanation of AI-informed decision-making goes towards the fundamental aim of ensuring transparency in government decision-making. According to Mashaw, to attain bureaucratic justice, decision-making systems should not only make accurate and cost-effective judgments, but also give attention to the dignity of the participants.[72] The dignitarian element means that those who are subject to an automated process should know or understand what reasons are behind a decision, rather than the system being an impenetrable 'black box'. As highlighted by Oswald, incorporating an algorithm into decision-making 'may come with the risk of creating 'substantial' or 'genuine' doubt as to why decisions were made and what conclusions were reached, both for the subject of the decision and the decision- maker themselves'.[73] This suggests that the government must use software that furnishes relevant evidence to support evaluation and auditing and allow for technical accountability due to the demand for transparency.[74]

We support proposal 7's requirement that a 'technical explanation of the AI-informed decision that can be assessed and validated by a person with relevant technical expertise', as it will enable experts to give proper advice on the AI *system.* AI systems are typically 'trained' through exposure to large datasets by machine learning models, and produce results based on the recognition of patterns.[75] The steps leading to an AI decision therefore are not translatable into logic and reasoning in the way that human-made decisions might be, and whilst not random, can

[72] Jerry L Mashaw, *Bureaucratic Justice* (Yale University Press, 1983) 49. This issue was also stressed by the UK Supreme Court in the case of *R* (*Osborn) v Parole Board* [2014] AC 1115. Allsop CJ also placed emphasis on dignity in his speech 'Values in Public Law' published in (2017) 91 *Australian Law Journal* 118.
[73] Marion Oswald, Algorithm-assisted decision-making in the public sector: framing the issues using administrative law rules governing discretionary power (2018) *Phil. Trans. R. Soc.* A 376, p. 5 <http://dx.doi.org/10.1098/rsta.2017.035>.
[74] Deven R Desai and Joshua A Kroll, 'Trust but Verify: A Guide to Algorithms and the Law' (2018) 31 *Harvard Journal of Law and Technology* 1, 44. For the Australian guidelines, see Australian Government, Department of Finance and Administration, Australian Government Information Management Office, *Automated Assistance in Administrative Decision-Making: Better Practice Guide* (2007) 45-9.
[75] Robert French, *Rationality and Reason in Administrative Law*, 4.

be difficult to access and understand.[76] This is what is known as a 'black box' system.[77] The complex nature of coding and algorithms means that experts would most likely be required to elucidate a system's information and processes leading to a decision.[78]

Australia's Existing Legislative Requirements for Reasons for Decisions

There are two key laws that impose on public sector decision-makers obligations to provide reasons for decisions, on request by an applicant. Both the *Administrative Decisions (Judicial Review) Act 1977* (Cth) ('ADJR Act') and the *Administrative Appeals Tribunal Act 1995* (Cth) require decision-makers to provide on request: a statement in writing setting out the findings on material questions of fact, referring to the evidence or other material on which those findings were based and giving the reasons for the decision.[79]

These requirements are explained in a set of guidelines published by the Administrative Review Council.[80] As summarised by Groves, a statement of reasons must 'do more than simply list evidence and state the decision reached'.[81] They must also provide an explanation of 'the logic or "intellectual process" by which evidence was used to reach the decision'.[82] Moreover, as stated in *Campbelltown City Council v Vegan,* 'where more than one conclusion is open, it will be necessary … to give some explanation of [the] preference for one conclusion over another'.[83]

However, it is unclear how the existing administrative law frameworks will adequately provide reasons for AI-informed decision-making, which necessitates amendments to the legislation or new legislation to clarify these principles.

We would also recommend the following specifically regarding the explainability of design/purchase and implementation of AI-based decisions:

[76] Ibid; Danielle Keats Citron, *Technological Due Process* (2008) 85(6) *Washington Law Review* 1249 at 1277.
[77] Karen Yeung, Council of Europe Committee of Experts on Human Rights Dimensions of Automated Data Processing and Different Forms of Artificial Intelligence, *A Study of Implications of Advanced Digital Technologies (Including AI Systems) for the Concept of Responsibility Within a Human Rights Framework* (2018) 26; Access Now, *Human Rights in the Age of Artificial Intelligence* (2018) ('*A Study of Implications of Advanced Digital* Technologies'); Danielle Keats Citron and Frank Pasquale. 'The Scored Society: Due Process for Automated Predictions'(2014) 89 *Washington Law Review* 1, 6.
[78] Danielle Keats Citron, 'Technological Due Process' (2008) 85(6) *Washington Law Review* 1249, 1284.
[79] *Administrative Decisions (Judicial Review) Act 1977* (Cth); s13; *Administrative Appeals Tribunal Act 1995* (Cth), s28.
[80] Administrative Review Council, Practical Guidelines for Preparing Statements of Reasons, November 2002 <www.arc.ag.gov.au/Documents/arcguidelinesnew.pdf>.
[81] Matthew Groves, 'Reviewing Reasons for Administrative Decisions: Wingfoot Australia Partners Pty Ltd v Kocak' (2013) 35 *Sydney Law Review* 627, 630, citing *Hill v Repatriation Commission* (2004) 39 AAR 103; *Preston v Secretary, Department of Family and Community Services* (2004) 39 AAR 177; *Civil Aviation Safety Authority v Central Aviation Pty Ltd* (2009) 253 ALR 263.
[82] Ibid, citing *Garrett v Nicholson* (1999) 21 WAR 226 [73].
[83] (2006) 67 NSWLR 372, 397.

Design/Purchase Phase

- Where complex decisions are automated, it will be necessary to ensure that the statement of reasons for the automated decision or outcome is captured.
- Where machine learning is utilised to automate decisions, it is required to accurately document the decision logic, including:
    - the principles behind the machine learning model;
    - training and testing processes; and
    - a statement of reasons is logged for all predictions or decisions at the point in time that they are made.

Additional considerations apply where technology is not developed in-house.

- Purchase of off the shelf program
    - Procurement guidelines should address issues including proprietary code and how it can be used consistently with transparency requirements.
    - There is a possible role for standards which address issues such as fitness for purpose, bias, transparency, explicability, and accountability.

- Contracted out
    - Where the technology is contracted out, it is vital to ensure the agency has rights of possession in relation to the information necessary to ensure transparency/explicability. The Office of the Australian Information Commissioner could potentially issue advisory guidelines on this.

Implementation Phase

- Departments and agencies should ensure appropriate transparency, including ensuring that the data associated with the implementation and operation of the automated technology is created, stored and retained appropriately.
- Departments and agencies should ensure that there are suitable processes in place to meet Freedom of Information requirements and any applicable requirement to provide reasons for decisions.
- If the software involves the making of a decision that affects an individual, departments and agencies should ensure that the implementation accords with requirements in relation to procedural fairness.

In summary, we echo Berndt Wirtz, Jan Weyerer and Caroline Geyer, who suggest bodies that utilise automated and AI systems must ensure explicability and transparency, particularly with

algorithms and AI that govern humans' lives, in order to minimise the pitfalls and maximise the fairness of such technologies.[84]

[84] Bernd W. Wirtz, Jan C. Weyerer & Carolin Geyer, 'Artificial Intelligence and the Public Sector—Applications and Challenges' (2018) 42(7) *International Journal of Public Administration* 596, 603.

Question B: Where a person is responsible for an AI-informed decision and the person does not provide a reasonable explanation for that decision, should Australian law impose a rebuttable presumption that the decision was not lawfully made?

Proposal 10: The Australian Government should introduce legislation that creates a rebuttable presumption that the legal person who deploys an AI-informed decision-making system is legally liable for the use of the system.

The Castan Centre understands that the Australian Human Rights Commission has defined an AI-informed decision-making system to pertain to instances where AI materially assists in the process of making a decision. Importantly, this definition does not include the possibility of an AI system actually forming and executing a decision. It instead views an AI system as a tool to inform an outcome. The definition adopted therefore seems to assume that AI systems operate as a program in which a user can control the final decision.

We are of the view that assigning legal liability to a user of an AI system should only apply in instances where the user has the ability to control the final decision. On this basis, we consider that the proposed legislation, and rebuttable presumption, should only apply to the narrow forms of AI systems considered by the Australian Human Rights Commission. It should not apply in instances where an AI system could adapt and generate an outcome not intended or reasonably foreseen by the user of the AI system.

Further, we consider that legal liability should only be assigned to a user in this narrowed context, when there is one user of the system, such as a government. It is conceivable that there could be instances where multiple parties use an AI system when making a decision, especially in a private law context. Assigning liability based on use alone, could therefore be complicated by a scenario where there are multiple users of the system.

We therefore do not think it will be sufficient to introduce legislation that the legal person who deploys an AI-informed decision-making system should be legally liable for the use of the system, unless the definition of such a system is sufficiently narrow.

Question C: Does Australian law need to be reformed to make it easier to assess the lawfulness of an AI-informed decision-making system, by providing better access to technical information used in AI-informed decision-making systems such as algorithms?

The Castan Centre is broadly supportive of the view that the Australian law should be reformed to increase the transparency of AI-informed decisions. We however query the utility in permitting access to highly technical aspects of these systems, such as the underlying algorithms.

Reviewing the technical aspects of these systems in order to assess their lawfulness would require a specific expertise set held by a minority of persons. Importantly this expertise may not be held by those affected by the system. Consequently, it is likely that the proposal will not have the practical effect of increasing decision-making transparency.

Permitting access to the technical aspects of a system is also likely to undermine the proprietary value of the system as a whole. This in turn could act as a disincentive to developers of such systems.

Despite this, we acknowledge however that there is a need under the rule of law for individuals affected by a decision – AI-informed or otherwise – to be able to access and understand how a decision affecting them is reached.

Balancing the need for practical transparency and the interests of developers is difficult. We are of the view that instead of reforming the law to enable access to the technical aspects of AI systems, consideration should be given to alternative methods of increasing transparency, in particular methods which champion the use of plain English. For example, developers of these AI systems could be required by law to provide a 'disclosure statement'[85]. This statement could be required to explain in non-technical terms key information, such as: an overview of the decision-making process, including how persons participate and how decisions are ultimately reached. The disclosure statement should not require developers to disclose the intricacies of the AI in use, but rather provide a non-technical overview of the reasoning adopted by the system.

A disclosure statement model would increase transparency around the decision-making processes of these systems. It would do so in a way that is practical and understandable to those affected by the system's decisions. Yet, this model would also protect the proprietary value of a system's technical aspects.

[85] Such a disclosure statement could be akin to the Product Disclosure Statements used in the financial services and products market. See in particular: *Corporations Act 2001* (Cth) ch 7, pt 7.9, div 2.

Question D: How should Australian law require or encourage the intervention by human decision makers in the process of AI-informed decision making?

The Castan Centre acknowledges that there is a difference between AI-informed decision-making and instances where an AI system itself makes a decision. That is, there is a difference between instances where a natural person relies upon information generated by an AI system in making a decision and where an AI system uses information generated to make a decision. We note that the latter form is outside the scope of this proposal, although it raises interesting questions.

In terms of AI-informed decision-making, we are of the view that immersing a natural person into the reasoning process may defeat the potential benefits of using AI in the first instance. For example, it holds the potential to reintroduce implicit bias into the system and to reduce the efficiency of decision-making processes.

Additionally, when AI-informed decisions are made in a governmental context, there is a practical issue of who should be authorised to make a decision. A minister may delegate their authority within the public service for the purposes of decision making. This in turn raises the question of to whom the minister should delegate their authority and what degree of seniority they should have, if a natural person is to intervene in such decisions.

In instances where AI is used to inform a decision, we do not support the notion that a natural person should be allowed at law to intervene in the reasoning process of an AI system. Instead, we are of the view that Australian law should enable natural persons to intervene post the use of an AI system but before a decision is made and executed. That is, Australian law should permit a natural person to review data generated by an AI system and determine whether the position recommended by the system should be adopted or overridden.

Here it is useful to consider a key mechanism in the European General Data Protection Regulation (GDPR). In particular, Article 22 requires data controllers uphold the rights of data subjects. The provision grants a data subject the '*right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision'*.

In terms of AI-informed decision-making processes, Australian law should be reformed to mirror this concept from the GDPR. That is, the law should entail three key aspects. Firstly, after an AI system has completed its reasoning and generated a recommended decision, a natural person should be required to review this recommendation. In reviewing the proposed option, the natural person should be required to take into consideration the human rights of an individual impacted by an AI-informed decision, before opting to adopt the recommendation.

Secondly, individuals impacted by an AI-informed decision should be provided with a mechanism to a) express their opinions or concerns about the AI-informed decision; b); request a review by a natural person of the decision that was made utilising the AI-informed data; and c) have an avenue to contest the decision.

In terms of implementing c), an appeals tribunal consisting of members with the necessary AI technical and legal expertise could be established to review contested AI-informed decisions made in a governmental context.

Question E: In relation to the proposed human rights impact assessment tool in Proposal 14

1. **When and how should it be deployed?**
2. **Should completion of a human rights impact assessment be mandatory, or incentivised in other ways?**
3. **What should the consequences be if the assessment indicates a high risk of human rights impact?**
4. **How should a human rights impact assessment be applied to AI-informed decision-making systems developed overseas?**

(a) Context

We prepared responses to the AHRC's questions about human rights impact assessments (HRIAs) based on a review of the sources listed in the table set out in Appendix 1. We focused our review on sources that make recommendations around HRIAs in the AI space in particular. The following recommendations are based on what appears to be a common theme from the sources reviewed:

(b) Recommendations

Q1 When should the HRIA be implemented

The HRIA should be implemented in all instances where AI systems or technology can make decisions that impact upon the human rights of potential subjects or groups. It should be utilised from development and acquisition of an AI system through to implementation and use with regular ongoing monitoring and evaluation by independent entities and experts.

Regular and ongoing review of the AI system to monitor its compliance with HR should be tailored to the system in question and follow that system's life cycle. Review should (at least) be undertaken at each new phase of the system, such as a new any changes to its use (e.g. roll-out, introduction of new technology et cetera). This will allow for a review process that takes account of the particularities of specific systems and is conducted to reflect on any important changes to the system, regulations and new tech introduced and applied to the system.

Q2 How should the HRIA be implemented

1) Preliminary steps

- Map existing HRIAs used/applied in other areas
- Consultation with academics, the tech industry, human rights groups, NGOs and the public.

2) Development of guidelines/checklists

- Develop guidelines and checklists for HRIAs of AI systems
    - These could be modelled off existing guidelines for HRIAs in other context but must be adapted to the AI space through consultations mentioned above
- Develop a template self-assessment form/checklist for entities to use when conducting HRIAs.

3) Self-assessment as per guidelines and templates

- Public authorities and private entities wishing to acquire, develop, use and implement an AI system should conduct a self-assessment as per the guidelines mentioned above and using a uniform template
- Assessments should include plans for mitigation and prevention of any risks identified.

4) Submit assessment to an independent review body

- This may be an existing body, new body within an existing body or a new body;
- The independent entity should review assessments and have oversight
- HRIA assessments should be made available to the public and a complaints process should be in place for assessments deemed to be inadequate by the review body or members of the public.

Q3 Should completion of a HRIA be mandatory, or incentivised in other ways?

The commentary in the table in Appendix 1 appears to reflect a prevailing view that HRIAs should be mandatory for all AI systems. Particular emphasis was placed on systems that presented a high risk to human rights, and systems that were being implemented by public authorities. From a human rights perspective however, the mandatory implementation of HRIAs for all private and public bodies notwithstanding high-risk status would facilitate uniformity in AI regulation, encourage greater transparency in the use of AI technologies, and enable more effective identification and management of systems that pose a high-risk to human rights.

Q4 What should consequences be if the HRIA indicates high risk of HR impact

The research indicates that high risk systems should be reviewed by an independent authority and their use suspended in the interim. This is particularly important where such systems have been procured by government bodies. The review process should include consultations with experts and affected groups in order to conduct a meaningful review of the human rights implications of using the AI technology in question. The outcome of these independent assessments should be made publicly accessible to ensure transparency.

Private and public bodies should be given the opportunity to provide technological rectification develop measures to prevent recurrence in the future, and remedy any harm already caused. In cases where adequate steps cannot be taken to implement such safeguards against human rights breaches, use of the systems must be discontinued.

Q5 How should HRIA be applied to AI systems developed overseas?

As for AI systems developed beyond Australian jurisdictions, the prevailing view indicates that the use of such systems must be conditional on the proper employment of HRIAs by public bodies in conjunction with overseas third party organisations. These third party organisations therefore would need to waive trade secrecy, confidentiality and other restrictions that impede the function of HRIAs. Where a third party organisation refuses to do so, procurement or use of the technology by public bodies should be discontinued. Public bodies should also maintain regular oversight of such systems throughout their use to ensure ongoing compliance with human rights protections.

Proposal 11: The Australian Government should introduce a legal moratorium on the use of facial recognition technology in decision making that has a legal, or similarly significant, effect for individuals, until an appropriate legal framework has been put in place. This legal framework should include robust protections for human rights and should be developed in consultation with expert bodies including the Australian Human Rights Commission and the Office of the Australian Information Commissioner

Context

Facial recognition technology is being increasingly used by law enforcement bodies both overseas and within Australia.[86] The technology has drawn significant criticism for producing erroneous facial matches, and in some cases has been used improperly in order to obtain matches to aid investigation.[87]

Facial recognition has led to litigation in overseas jurisdictions. For instance, in August 2020, Sweden's data protection authority fined a local agency more than $20,000 for a three-week test of a facial recognition system that logged each time a student entered a classroom. It was the country's first enforcement action under GDPR.[88] In October 2020, France's data protection regulator said schools should not use facial recognition to control who entered, after receiving complaints about plans to test the technology at high schools in Nice and Marseilles.[89]

The AHRC's Discussion Paper mentions that 'several American jurisdictions, for example, have passed or are considering laws banning use of facial recognition software where there is potential for harm'. Our research lists the jurisdictions that have some form of ban in place or are considering a ban on facial recognition software (see Appendix 2 Table of sources). We only found American jurisdictions (municipal and State level) with bans in place. The bans in place in the US tended to be bans (through ordinances/policy) at the municipal levels and moratoriums at the state levels (with some exceptions). The basis upon which American jurisdictions have enacted bans or moratoriums were largely similar, namely that the negative impact on human rights outweigh any purported benefits of the software.

We also note that a number of UK governmental reports have raised serious concerns with facial recognition:

- The UK Parliament's Science and Technology Committee states that any rollout of facial recognition technology beyond current pilots should be paused until concerns regarding bias and effectiveness 'have been fully resolved'.[90]

---

[86] Jon Schuppe, 'How Facial Recognition Became a Routine Policing Tool in America', *NBC News* (online), 11 May 2019 < https://www.nbcnews.com/news/us-news/how-facial-recognition-became-routine-policing-tool-america-n1004251>.

[87] Claire Garvie, 'Garbage In, Garbage Out: Face Recognition on Flawed Data' on Georgetown Law, *Centre on Privacy and Technology,* 16 May 2019 < https://www.flawedfacedata.com/#>.

[88] https://www.datainspektionen.se/nyheter/2019/facial-recognition-in-school-renders-swedens-first-gdpr-fine/

[89] https://www.cnil.fr/fr/experimentation-de-la-reconnaissance-faciale-dans-deux-lycees-la-cnil-precise-sa-position?

[90] Science and Technology Committee of the United Kingdom Parliament, Biometric strategy and forensics services (Fifth Report of Session 2017-19) (HC 800) (25 May 2018), 4.

- the UK Biometrics Commissioner and Forensic Science Regulator have recommended a moratorium on the use of facial recognition technology 'until a legislative framework has been introduced and guidance on trial protocols, and an oversight and evaluation system, has been established'.[91]
- The UK Equality and Human Rights Commission have also made a similar call:

  'In light of evidence regarding their inaccuracy and potentially discriminatory impacts, suspend the use of automated facial recognition and predictive programmes in policing, pending completion of the above independent impact assessments and consultation process, and the adoption of appropriate mitigating action'.[92]

Similarly, regional bodies and NGOs have called for a moratorium.[93] Liberty UK, for instance, have stated that:

> Facial recognition is a dangerously intrusive and discriminatory technology that destroys our privacy rights and forces people to change their behaviour. It has no place on the streets of a free, rights-respecting democracy.[94]

Given the US experience, and the arguments raised by NGO's and others, it appears strongly arguable that if facial recognition technology is not trained on diverse datasets, or is misused, certain racial groups will likely face more frequent misidentification and be subject to increased police scrutiny. This will likely constitute a form of discrimination that anti-discrimination law in Australia will need to contend with.

In terms of human rights impacts, we note that using facial recognition technologies to process facial images captured by video cameras in public space may interfere with a person's freedom of opinion and expression. In particular, where a person is under surveillance in certain contexts, they may fear potential consequences from participating in lawful democratic processes such as protests and meetings with individuals or organisations, including increased surveillance or scrutiny by police. This engages the right to freedom of association and assembly and freedom of expression and opinion.[95] In addition to this, it may potentially engage the right to be free from unlawful and arbitrary arrest.

---

[91] House of Commons Science and Technology Committee (United Kingdom Parliament) *The work of the Biometrics Commissioner and the Forensic Science Regulator* (House of Commons Paper 1970, Nineteenth Report of Session 2017-19, 18 July 2019), 4 <https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/1970/197003.htm#_idTextAnchor000>.

[92] UK Equality and Human Rights Commission, Civil and political rights in Great Britain — March 2020, at 89 https://www.equalityhumanrights.com/sites/default/files/civil_and_political_rights_in_great_britain_2020.pdf

[93] See, for example, https://thepublicvoice.org/ban-facial-recognition/endorsement/.

[94] https://www.libertyhumanrights.org.uk/resist-facial-recognition

[95] On this issue, se European Commission for Democracy through Law (Venice Commission) OSCE Office For Democratic Institutions and Human Rights (OSCE/ODIHR) *Guidelines on Freedom of Peaceful Assembly* (3rd ed) CDL-AD(2019)017; Strasbourg / Warsaw, 8 July 2019; p 25 [71] https://www.venice.coe.int/webforms/documents/?pdf=CDL-AD(2019)017-e

Recommendations

Based on the many potential harms and risks to human rights through the use of facial recognition software, and the lack of evidence of substantive benefits, it appears that on balance, a temporary moratorium should be imposed until adequate safeguards and detailed regulation has been put in place to prevent and mitigate against any risks to human rights.

We thank you for accepting our submission and look forward to engaging with the Commission on these issues in consultations.

Dr Maria O'Sullivan, Deputy Director, Castan Centre for Human Rights Law, Monash University, on behalf of the Centre's research team.

APPENDIX 1

Table of sources on the use of Human Rights Impact Assessments

Source	When should the HRIA be employed?	How should the HRIA be employed?	Should completion of a HRIA be mandatory, or incentivised in other ways?	What should consequences be if the HRIA indicates high risk of HR impact?	How should HRIA be applied to AI systems developed overseas?
Council of Europe Human Rights Commissioner Recommendations (2019)	When AI systems are acquired, developed or deployed by public authorities. From development to implementation. Review of the system should be undertaken on an ongoing and regular basis, at least at every new phase of the AI system.	Implementation should be similar to other regulatory impact assessments, such as data protection impact assessments. The public must be given access to research from HRIAs. Public authorities planning to 'acquire, develop or deploy' a system should conduct self-assessments. The self-assessment should account for the nature,	Self-assessment should be carried out before acquiring/ developing a new AI system. External reviews of systems should be conducted to measure impact over time.	If the review or self-assessment discloses a high risk of HR impact, the HRIA should set out safeguards and other measures to mitigate or prevent such risks from materialising. If high risk is identified in relation to a system that is already in operation, the system should be suspended until a plan for prevention and mitigation is put in place. If it is not possible to put a meaningful plan in place to prevent or mitigate against the	An AI system must only be obtained from a third party if that party is willing to waive any restrictions that impede a HRIA and making such HRIA publicly available.

		context, scope and purpose of the AI system. External reviews of the system in question should be undertaken by an independent entity or researcher to discover, measure or map HR impacts and risks over time. Public authorities should consider involving NHRIs to carry out this function. It should include an evaluation of how decision-makers collect or influence inputs and interpret outputs of AI systems. ..		risk, the AI system should not be adopted by any public authority. If a HR violation is found, public authorities must act to address and remedy the violation and take measures to prevent or remedy risks of it occurring in the future.	
Council of Europe Committee of Ministers	When algorithmic system has potentially significant HR impacts.	Consultation with affected and potentially affected individuals and groups should be conducted	HRIAs should be conducted for all systems with potential significant impact on human rights	Algorithmic systems with a high risk to human rights should include a HRIA that identifies possible transformations on social, institutional or government structures and	States and 'any private actors engaged to work with them or on their behalf', should conduct HRIAs

Draft Recommendations (2019)	At any stage of the lifecycle of the system. Review should be ongoing and regular during the lifecycle of the project.	Staff engaged in connection with an algorithmic system must be trained regarding relevant human rights and non-discrimination standards and human rights compliance.	HRIAs should be mandatory for systems with high risks to human rights. A 'computational experimentation' should only be used if a HRIA has been conducted. Algorithmic systems should not be used if confidentiality or trade secrets prevent a meaningful HRIA from being conducted.	include recommendations to prevent and mitigate the high risks. All HRIAs for systems with a high risk of adverse impact on human rights should be subject to an independent review and be publicly accessible if conducted for a public authority and include expert input and a follow-up mechanism. It may include trials before official release of the system to ensure that groups and individuals that may be affected are consulted and can participate in the decision-making process, design, testing and review phases. If prevention or mitigation against high risks of adverse human rights impacts is not possible, the system should not be used by public authorities. If the system is already being deployed, it should be suspended	before public procurement at all stages of the system lifecycle. A system should not be procured if confidentiality or trade secrets mean that a meaningful HRIA cannot be conducted. If private entities provide services that rely on algorithmic systems and the service is considered essential in modern society for effective enjoyment of HR, the State should preserve future viability of alternative solutions and continued access to such services by

				until measures of mitigation has been put in place. HR violations in connection with a system already in use should be identified and remedied immediately and measures put in place to prevent such violations in future. Any risks and harms should be addressed in a timely and adequate manner. Responses from private actors should be evaluated for effectiveness to prevent or mitigate adverse HR impacts.	affected individuals and groups.
UN Special Rapporteur on Freedom of Expression **Report (2018)**	From conception to implementation of AI systems.			Individuals must have remedies for any adverse human rights impacts that AI systems may have.	

AI Now Institute, **Algorithmic Impact Assessments: Practical Framework for Public Agency Accountability (2018)**	Review of impacts should be monitored overtime through an external review process and audits.	Automated decision systems should be assessed by public authorities through a self-assessment to evaluate potential impacts on fairness, justice, bias or other concerns across affected communities. Review and research of systems should be made publicly available and the public should be consulted to clarify questions and concerns.	Public authorities should conduct a self-assessment of existing/proposed automated decision systems evaluating impacts on fairness, justice, bias and other concerns on affected communities	Individuals and groups should be able to challenge inadequate assessments and system uses that the public agencies have failed to mitigate, correct or prevent.	Trade secrets should not bar meaningful research on automated decision systems. Public agencies seeking to acquire a system may need to ask potential third parties to waive restrictions on information to enable external research and review of the system. Vendors should at least be contractually required to wave any
The Human Rights, Big Data and Technology Project (University of Essex)	At the start of any project by States and businesses, followed by ongoing and regular monitoring and evaluation.	The HRBAs would be strongest if carried out by 'independent participation and oversight', such as through a combination of parliamentary committees,	States and businesses should conduct HRBAs of all current uses of big data and AI.	If any impact to HR is identified, the State/business should take steps to end negative effects. For example, redesigning the relevant algorithm or removing automation from the decision-making process.	

UDHR at 70: Putting Human Rights at the Heart of Design, Development, and Deployment of AI (2018)		judicial/quasi-judicial bodies and courts		The right to a remedy is separate from technological rectification. The right to a remedy includes, prevention, redress and non-occurrence in future. It is therefore not solely reactive in face of actual violations but preventative.	
Toronto Declaration: Protecting the right to equality and non-discrimination in machine learning systems (2018)	Before development and acquisition of the machine learning system and where possible before use. An ongoing investigation must be undertaken during the system's lifecycle.	Regular impact assessments to identify potential sources of discriminatory or other HR harms, for example in the algorithmic design, oversight process or processing Take measures to mitigate risks identified, for example by conducting testing, pre-release trials, include potentially affected groups	Any State deploying machine learning systems must 'thoroughly investigate systems for discrimination and other HR risks before development or acquisition and where possible, prior to use.	States should refrain from using 'black box systems' that are not susceptible to meaningful transparency and accountability in high-risk contexts.	States should maintain oversight and control over a system procured by a third party and require third parties to conduct human rights due diligence to identify, prevent and mitigate discrimination and HR impacts and disclose these steps publicly.

		and experts in the decision-making process. Subject systems to regular and live tests and audits, check markers of success for bias and self-fulfilling feedback loops, ensure independent reviews of the systems in a live environment. Disclose system limitations, for example failure scenarios. The use of machine learning should be disclosed publicly, including action taken to mitigate against harmful impacts. Independent analysis and oversight of the system that is audited must in place.			Private sector developing machine learning systems should follow the human rights due diligence framework to avoid violation rights: - Identify HR impacts - Take steps to prevent and mitigate against impacts and monitor responses - Be transparent about efforts taken

		Independent oversight, including by judicial authorities should be available where necessary. Public bodies should carry out training in human rights and data analysis for officials involved in procurement, development, use and review of machine learning tools.			

APPENDIX 2

Table of sources on facial recognition moratorium

Jurisdiction	Scope of ban/ moratorium	Basis
US Municipal level		
San Francisco (May 2019) *Ban (ordinance)*	- Local agencies (e.g. police, transport authority, law enforcement) using FR - Buying FR software requires approval from city administrators - Excludes: Airports and seaport as run by federal agencies - Amended (Dec 19) to allow Apple FaceID and similar tech products with facial recognition if necessary to the job and no viable alternatives	- The propensity of facial recognition technology outweighs purported benefits - Surveillance historically used to oppress minorities - If and how tech should be funded, acquired, used, shared requires meaningful public input - Legally enforceable safeguards (including transparency, oversight, accountability measures) must be in place to protect human rights before tech deployed - Reporting measures must be adopted to verify human rights adhered to

Somerville **(June 2019)** *Ban (ordinance)*	- Any department, agency, bureau, and/or subordinate division of the City and any person/entity acting for the city obtain, retain, access, use FR or information obtained from such	- Benefits are few and speculative and greatly outweighed by its substantial harms - Broad application in public spaces functional equivalent of requiring every person to display personal identification at all times - Far less accurate in identifying women, young people, people of colour – elevated risk of harmful false positive - Many databases plagued by racial and other biases, generating 'copycat' biases - Public use can chill free speech - Broad application in public spaces functional equivalent to requiring all persons to carry identification at all times

Oakland **(July 2019)** *Ban (ordinance)*	- Any department, agency, bureau, and/or subordinate division of the City acquiring, obtaining, retaining, requesting, accessing FR tech - Requiring city staff members to obtain approval from chair of Oakland's Privacy Advisory Commission before 'seeking or soliciting funds' for FR tech. State and federal funding must also be approved - The City shall only approve an action if first considering rec from the Privacy Advisory Commission, and whether tech outweigh the costs, will safeguard HR, and in the City's judgment no alternative with lesser economic cost or impact on HR as effective	Not expressed in the copy of the ordinance. Police not currently using FR tech but now prevented from doing so.
Berkeley **(October 2019)** *Ban (ordinance)*	- Approval from City Manager required (except in exigent circumstances) to seek, solicit, accept funds, acquire tech, use tech, enter an agreement with non-City entity to acquire, share, use tech or info - Excludes a long list of devices, such as: routine office hardware; handheld parking devices; manually-	- Council members noted the potential use of FR on a broad scale to track people would be an 'egregious violation of the Fourth Amendment'

	operated portable digital cameras, auto recorders, video recorders (not to be used remotely, function limited to manually capturing, viewing, editing, downloading video	
Alameda **(December 2019)** *Ban (through policy)*	- Policy to ban the use of FR tech. Council called n staff to create a binding ordinance to ban future use of FR tech together with data protection and privacy oversight ordinances	Unable to obtain copy of policy.
Brookline **(December 2019)** *Ban (ordinance)*	- Town voted to ban government use of FR tech at a town meeting	Unable to obtain further information around ban.

Northampton (December 2019) *Moratorium (ordinance)*	- Council banned the government from collecting and using biometric information through surveillance tech - Any city official to expend any city resources to obtain, retain, access or use face surveillance system - After three years, the ordinance shall be reviewed	- Council President expressed concern that tech is outpacing regulation and state legislation prohibiting tech largely non-existent. Also concern over human bias in the tech.
Cambridge **(January 2020)** *Ban (ordinance)*	- City Manager must seek approval from the City Council before seeking funds, acquiring, using or entering into an agreement to acquire, share or otherwise use surveillance tech. - All departments (except police department) must submit an impact report and if necessary a use policy re specific tech for which approval sought - If police department (and unless an exigent circumstances), impact report and if necessary, a policy, before acquiring tech - In 'exigent circumstances' the police may acquire tech not exceeding 90 days without other procedures, must	- Councillor stated city is stepping up to protect residents from intrusive and undemocratic tech - To safeguard right to privacy, and balance privacy with need to promote and ensure safety and security - Provide protocol for use of surveillance tech with assessment of costs and protection of HR of individuals and communities (including communities of colour, and other marginalised communities) and allow for informed public discussion before use - Provide transparency, oversight, accountability and minimise risks

	report acquisition within 90 days of the end of the period and submit an impact report and if necessary use policy, an include in annual surveillance report. Extensions may be granted. - The ordinance includes a number of exemptions and exceptions, such as: data obtained when individual voluntarily and knowingly consented to information (i.e. receipt of city services), opportunity to opt out, pursuant to warrant. - Includes whistleblower protection for city staff that report a violation	
US State level		
California **(October 2019)** *3-year moratorium*	- Police departments and law enforcements prohibited for 3 years from directly using biometric surveillance systems, request or agree that another agency or third party use such system on their behalf or installing, activating, using such system relating to an officer camera or data collected by officer camera	- Californians value privacy as an essential element of individual freedom and guaranteed under the Californian Constitution - FR poses unique and significant threats to HR

	- Breaching the moratorium may result in sanctions and penalties and a person subjected to it may bring an action for equitable or declaratory relief in court - Not included: Use of mobile fingerprint scanning during lawful detention to identify a person without proof of ID if does not generate in retention of biometric data.	- FR functional equivalent to require every person to carry identification at all times in violation of constitutional rights. - FR allows people to be tracked without consent and generate massive databases about law-abiding citizens and may chill free speech in public places - FR repeatedly demonstrated to misidentify women, young people and people of colour, risking high level of false positives - Corrupt core purpose of officer-worn body-worn cameras by transforming these from transparency and accountability tolls to roving surveillance systems - Disproportionately impact on HR of those in highly policed communities and diminish effect of policing and public safety by discouraging people in these communities, including victims of crime, undocumented persons and people with unpaid fines and fees and with prior criminal history from seeking police assistance from the police
Massachusetts **(Ongoing)**	- Unlawful (unless express statutory authorisation) for any government official or branch of the Commonwealth of Massachusetts or any authority	- Racial disparities in databases used for facial recognition technology

Proposed moratorium (Bill)	established by the general court to serve a public purpose to acquire, possess, access or use biometric surveillance system or information obtain from such.
	
	- Statutory authorisation for use must include:
	  - Entities permitted to use the systems, purpose of the use and prohibited uses
	  - Standards for use and management of information derived from the system including data retention, sharing, access, audit trails.
	  - Audit for accuracy rates by gender, skin colour and age
	  - Protection for due process, privacy, free speech and association, and racial, gender and religious equity
	  - Compliance mechanisms
	  - Information obtained contrary to the legislation should not be admissible by government in any criminal, civil, administrative or other proceeding
		- Less accurate in identifying women, young persons and people of colour
		
		- Broad application akin to requesting all persons to display a personal photo identification card at all times, a mass violation of privacy
		
		- Application may chill exercise of free speech and association
		
		- Benefits outweighed by substantial harms

Michigan **(Ongoing)** *Proposed ban (Bill)*	- Any law enforcement official banned from obtaining, gaining access to or using real-time FR tech or information obtained from such when enforcing state laws.	Details not included in the Bill.

	- Evidence obtained in breach of the moratorium would be excluded as if it were in violation of the Constitution - Exception: The ban would not apply to the use of real-time tech under a belief that an emergency existed involving imminent risk to an individual(s) death, serious physical injury, sexual abuse, live-streamed sexual exploitation, kidnapping, human trafficking and the use of the real-time FR tech could prevent or stop the emergency	
Michigan **(Ongoing)** *Proposed 5-year moratorium (Bill)*	- Any law enforcement official (including police) would be prohibited for 5 years to obtain, access or use FR tech or information obtained from such tech to enforce state laws. - Evidence obtained in breach of the moratorium would be excluded as if it were in violation of the Constitution	Details not included in the Bill.
New Hampshire **(Ongoing)**	- Any department, agency, bureau, or administrative unit of the state of New Hampshire (including any city, town, or municipal entity) and any person or entity acting on behalf of the state are prohibited from	Details not included in the Bill.

Proposed ban (Bill)	obtaining, retaining, accessing or using facial surveillance system or any information obtain from such - Evidence obtained in breach of the ban would be inadmissible - A person aggrieved by violation of the ban has a private cause of action for injunctive relief, declaratory relief or writ of mandamus in any court and entitled to recover damages - A violation by a state employee would result in consequences that may include retraining, suspension or termination	
New York **(Ongoing)** *Proposed 2-year moratorium (Bill)*	- A proposed moratorium on the purchasing and use of biometric identification tech (including FR) in schools - Review by Commissioner and Department's Chief Privacy Officer to make recs to government whether biometric identifying tech including but not limited to FR, appropriate for use in public and non-public elementary and secondary schools, including charter	Details not included in the Bill.

	schools and if so, what restrictions and guidelines should be enacted to protect HR	
New York **(Ongoing)** *Proposed ban (Bill)*	- Any police agencies would be banned from using FR tech	Unable to locate Bill.
Oregon **(Ongoing)** *Proposed ban* *(Bill)*	- FBI and ICE would be banned from access to the State's driver's records for FR identity checks in the context of attempts to find and potentially deport people without proof of citizenship or legal residence.	Unable to locate Bill.

Utah **(Future)** *Potential legislative action*	- The legislature is considering action after revelation of the use of FR tech by federal law enforcement to scan driver's licence photos - Calls for regulation of the use by the Department of Public Safety of FR tech	
Washington **(Ongoing)** *Proposed 3-year moratorium (Bill)* *Note: other similar Bills*	- It would impose the moratorium on the operation, installation or commission of such FR by any person in any place of public resort, accommodation, assemblage or amusement - It would be unlawful for any Washington state or local government agency or official to obtain, retain, request, access or use FR tech or information obtained from or by use of such - Inadvertent or unintentional receipt, access, or use of information obtained from FR tech not a violation if it was not requested or solicited by agency/official, information permanently deleted upon discovery	Details not contained in the Bill.

also considered: here and here	- Evidence obtained in breach of the moratorium would be inadmissible - Person may institute proceedings for breach of the moratorium to claim injunctive relief, declaratory relief or write of mandate - Entitled to recover damages if subjected to FR - The moratorium would not apply to the Department of Licensing - It would establish a joint legislative task force on FR tech	
Non-US jurisdictions		
EU	- A Draft White Paper (obtained by Politico) put forward different options for regulation, including a moratorium 3-5 years.	

(Under consideration)	- A moratorium was dropped in the final version of its White Paper on AI published on 19 February.	