Australian Government
Australian Digital Health Agency

Australian Digital Health Agency submission to the OAIC's Consultation on Guidance for Transparency in Automated Decision Making

As emerging technologies shift to become embedded across various sectors of government and the economy, it is timely to consider the balance between the use and governance of automated decision-making (ADM) and the level of transparency for those who may be impacted by ADM. As an APP entity, the Australian Digital Health Agency (the Agency) supports the OAIC in development of the Guidance for the ADM Transparency Obligation and has prepared the following responses to questions in the consultation paper, contemplating ADM and transparency from several perspectives including legal, architectural and cyber.

Substantially and directly related to making a decision

Q.1	**In your view, what are the relevant factors to enable an entity to assess whether a computer program substantially facilitates and is directly connected to a human decision-maker's decision?** **For example:** • **Degree of reliance on the ADM system output** • **Ability and likelihood of human override over an ADM decision** • **Nature of the output (advisory vs determinative)** • **Transparency and explainability of outputs** • **Integration of ADM into decision-making workflow** **Please rank the factors from 1- 5 and include any additional factors.**

The identification of relevant factors to assess whether a computer program substantially facilitates and is directly connected to a human decision-maker's decision is dependent on the context of a given scenario and requires consideration of technical, control, and behavioural factors. All factors listed above are relevant to an assessment, and ranking in terms of relevance would be appropriate to consider when presented with a particular scenario.

If considering the listed factors from a standalone perspective however, the following ranking could be applied:

1. Nature of output (advisory v determinative)
2. Degree of reliance on ADM system output
3. Transparency and explainability of outputs
4. Ability and likelihood of human override over an ADM
5. Integration of ADM info decision-making workflow

Other relevant factors that may be considered include:

- The extent of reliance on system outputs – Role of the computer program in decision-making and the extent to which the computer program is used to inform, guide, or assist the decision-maker in exercising their discretion, or making judgement calls on which factors are relevant and irrelevant to the making of the decision.
- The determinative nature of the system's recommendations – How does the computer program calculate and arrive at a particular recommendation.
- Whether decision-makers can meaningfully override outputs – Is there a means for decision-makers to stand aside a recommendation or output when necessary to do so. Where reliance is high and override is limited in practice, the system is more likely to be considered directly connected to the decision.
- The degree of integration into decision workflows – In particular, specificity of the decision-making power and whether the power is conferred to a particular decision-maker (for example, a power that must be exercised personally by a Minister), or a general class of individual.
- The level of transparency – The degree to which the decision relates to a person and requires personal information for the ADM to provide an output, and the extent to which it is known by the person that the computer program is connected to a decision involving their personal information.

Fictional Edge Case: Generative AI with human oversight
The National Aged-Care Insurance Agency (NACIA) makes decisions about eligibility for elderly people trying to access life-changing aged-care support. To make eligibility assessments quicker, NACIA staff use a generative AI chatbot, GPTea, to summarise candidate profiles and receive recommendations about eligibility. NACIA's human staff always have oversight and make the final decision about eligibility based on the recommendation provided by the chatbot.

Q.2	**Would you consider NACIA's described uses of GPTea to be substantially and directly related to making decisions? Yes/No**
Yes.	

Q.3	**Can you provide your reasons for your response?**

The uses of GPTea in this scenario are substantially and directly related to making a decision as GPTea's summaries and recommendations are likely to materially shape and influence eligibility decisions, even when humans retain final authority. While the scenario highlights that AI summarises candidate profiles, it also goes further by providing recommendations about eligibility. The provision of recommendations necessarily involves a degree of analysis and reasoning to arrive at a suggested outcome, and therefore falls within the ambit of decision-making.

Notably, page 12 of the consultation paper states that the phrase *"do a thing that is substantially and directly related to making a decision"* reflects that a computer program may be used to recommend a decision to a human decision-maker.

Although human staff are said to retain oversight and make the final decision, there are material risks associated with this model. In particular:

- decision-makers may rely on AI-generated summaries, which may be inaccurate or contain hallucinatory information;

- automation bias may arise, leading decision-makers to give undue weight to AI outputs; and
- the use and influence of AI may expand over time.

These risks are likely to compound as decision-makers become accustomed to using computer programs and their outputs to justify decisions and increasingly rely on them in practice.

Meaning of significantly affect rights or interests

Q.4	**What factors do you consider increase the likelihood that a decision could affect an individual's rights or interests?** **For example:** • **sensitive information** • **vulnerable persons** • **intrusive practices** • **financial outcomes** • **other.** **Please rank using the above mentioned ranking system and list any additional factors.**

All factors listed above would impact the likelihood that a decision could affect an individual's rights or interests – likelihood increases where decisions involve sensitive data, vulnerable individuals, or materially affect access to services or financial outcomes, particularly where processes are intrusive, opaque, or difficult to challenge.

With regards to ranking the above, the Agency considers the following:

1. Intrusive practices
2. Vulnerable persons
3. Sensitive information
4. Financial outcomes

Other relevant factors that may be considered include:

- Conferral or removal of benefits in terms of whether the decision results in a benefit being granted to an individual, or an existing benefit being reduced or withdrawn.
- Impact on liberty and participation in society to the extent to which the decision affects an individual's liberty, freedom of movement, or ability to participate in society (for example, decisions relating to migration detention).
- Complexity and legal character of the decision – the complexity of the decision, including the legal framework in which it operates, and the degree of legal judgment required.
- Differential treatment and the extent to which the decision results in an individual being treated differently from others in comparable circumstances.

The Agency believe the *Privacy Act 1988* could also offer assistance with identifying factors that increase the likelihood that a decision could affect an individual's rights or interests by reviewing the definitions and inclusions for 'personal information', 'sensitive information' and 'health information' – in digital health and wider healthcare, this could include information such as decisions concerning or affecting children, genetic and biometric information, specialist reports and test results, and unique identifiers.

Q.5	**What classes of person do you consider to be vulnerable for the purposes of considering a decision's expected effect on a person's rights or interests, and why?**

Vulnerability may increase the likelihood that a decision will adversely affect an individual's rights or interests due to:

- Historical disadvantage and discrimination with the impact of past and, in some cases, ongoing discrimination in Australian society and law, potentially compounding the effects of current decisions.
- Intergenerational impacts and the effects of historical government decision-making, including policies that have resulted in intergenerational disadvantage or trauma.
- Barriers to participation and self-advocacy – limitations on an individual's ability to understand, engage with, or challenge decision-making processes. For example, some children, older persons, or people with certain disabilities may rely on others for support and advocacy.
- Limited inclusion in codesign and the extent to which affected groups have not been adequately involved in the design of laws, policies or decision-making systems (including automated systems) affecting them, increasing the risk of disproportionate or unintended impacts.

Classes of persons who may be considered vulnerable for these purposes include:

- minority groups (including religious, cultural and linguistic minorities)
- Aboriginal and Torres Strait Islander peoples
- older persons
- people living with disability or impaired decision-making capacity
- children and young people
- refugees and asylum seekers
- victims of domestic and family violence
- LGBTQIA+ people
- people with health conditions or care/support needs
- people experiencing socio-economic disadvantage
- people with low literacy, digital literacy, or language barriers
- people at risk of isolation, exploitation, or dependency

The basis for vulnerability will vary both between and within these groups.
Consideration should also be given to general vulnerability of individuals who do not have an alternative to providing their personal information into an ADM process i.e. if a person has no choice but to submit personal information to ADM in order to interact with government services, the alternative would likely be they lose access to government supports.

Q.6	**What do you consider to be a significant service or support, for the purposes of considering whether a decision affects access to a significant service or support?** **For example:** • **physical assistance/support** • **financial assistance** • **access to educational services 'in home services'**

	• **access to essential banking and credit services** • **access to telecommunications** • **access to essential utilities** • **Other.** **Please rank using the above mentioned ranking system and add any additional factors.**
The significance of a service or support should be assessed based on its necessity to basic needs, the individual's level of dependence on it, and the extent to which adverse decisions would materially affect their wellbeing, autonomy, or ability. All of the factors listed above increase the likelihood that a decision may affect an individual's access to a significant service or support. With regards to ranking the above, the Agency considers the following: 1. Physical assistance and support 2. Access to essential utilities 3. Access to banking and credit services 4. Financial assistance 5. Access to telecommunications 6. Access to educational services 'in home services'. Other significant services or supports should include: • health services and provision of care in health settings • disability supports, including the NDIS • government-funded pensions, allowances, and benefits	

Q.7	**Are there any other legal frameworks or policies that the OAIC should have regard to in considering the meaning of 'rights' and 'interest'?**
The OAIC could consider privacy, administrative, human rights, anti-discrimination, and AI governance frameworks, as these collectively define and protect the rights and interests affected by decision-making. It is also recommended that the OAIC consider whether there is a settled meaning of 'rights' and 'interest' in case law and common law. In addition, APP entities will need to consider any laws specifically applicable to their organisation when adopting and using ADM, including the technical build and operation of ADM systems and their functionality. This includes organisation- or system-specific legislation, such as the *My Health Records Act 2012.* OAIC guidance on ensuring ADM is compliant with specific legislation is welcomed.	

Fictional Edge case: Differential pricing

E-commerce company, Daintree, allows subscribers to order books to their home. To make an account, Marco was required to disclose his address, being the luxurious Capital Crescent in Sydney CBD. Marco is surprised that his desired book, The Art of Inheritance, was priced at $100 on Daintree when his book club buddy, Charles, from a country town, paid only $35 from Daintree. Marco is shocked to read an article reporting that Daintree uses differential pricing to inflate prices for individuals from wealthy postcodes. Marco is a freelance artist, so he does not have enough disposable income to purchase the book.

Q.8	**Restricting access to financial products or healthcare would clearly be a decision which significantly affects the interests of an individual. In the above Edge Case, would you**

	consider the inflation of prices for a consumer product to have significantly affected the interests of Marco? **If so, at what amount would differential pricing be considered to have significantly affected the interests of an individual? For example, 5% price difference, 20% price difference or higher.**
N/A. The Agency has elected not to provide a response to this question, due to the nature of our operation as a government entity stewarding national digital health infrastructure and products. We encourage the OAIC to take guidance on this question from other entities which undertake retail or financial operation.	

Q.9	**In judging whether differential pricing significantly affects rights or interests, how important would you consider the following factors on a scale from 1 to 6 (with 1 being not at all important and 6 being very important):** **a. the dollar amount of differential pricing** **b. the significance of the product or service being purchased?**
N/A. The Agency has elected not to provide a response to this question, due to the nature of our operation as a government entity stewarding national digital health infrastructure and products. We encourage the OAIC to take guidance on this question from other entities which undertake retail or financial operation.	

<u>Meaning of making a decision</u>

Fictional Edge case: Discriminatory targeted job ad
Engineering firm, Manscon, is hiring for graduate engineers. To try to target the best candidates, Manscon advertises open roles on job site, Opaque Window. Opaque Window algorithmically promotes this job posting to candidates it thinks are best suited. One metric it uses is gender of account holders. Since most engineers at Manscon are male, Opaque Window prioritises promoting the job to male engineering graduates. Rafqa is a female engineering graduate searching for a job on Opaque Window. She does not receive the Manscon job ad, which limits her employment options.

Q.10	**Should it be considered a 'decision' that Rafqa did not receive the Manscon job ad?**
Yes, this constitutes a decision, as the computer program is effectively filtering job postings on the basis of gender. Implicit to this functionality is the exclusion of female candidates from the opportunity to be informed that Manscon is hiring graduate engineers. Whether the gender-based filtering arises from a design flaw, or an intentional setting, the computer program should not be determinative. This is because the operation of the program results in the imposition of a hidden condition on who may apply for a job (i.e. only male applicants that are made aware of the job opportunity). In turn, this limits the pool of potential applicants and affects who may apply for, and ultimately be offered, employment.	

In decision-making terms, the computer program is undertaking a form of pre-selection of candidates by restricting visibility of the job advertisement to male candidates. This amounts to differential treatment at a critical preliminary stage of the decision-making process, which directly affects access to employment opportunities. Further, the computer program pre-selecting male candidates in the way described is inconsistent with sex discrimination laws and will likely be unlawful. Overall, the use of the computer program in the way described in the scenario is not merely administrative or facilitative. It is substantively connected to decision-making, as it determines which individuals are given the opportunity to participate in the recruitment process, and therefore materially influences the ultimate employment outcome. In situations such as these, accountability thus belongs to the person who approved the implementation of the system especially since this outcome was readily foreseeable.

Meaning of 'arranged for'

Q.11	**Are there any scenarios relevant to the meaning of 'arranged for' that you consider require guidance? If so, why?**
The discussion of "arranged for" on page 20-21 of the consultation paper requires further guidance in general. It would be useful to have guidance on scenarios where an entity has arranged a computer program, to make parts of a decision, and also operates ADM for other parts. It would also be useful to have guidance on where an entity has arranged for a computer program to make a decision, versus an entity simply accessing or using ADM, in the context of inter-agency use of computer programs. This would be particularly relevant in the digital health ecosystem, where healthcare providers or organisations may utilise a computer system provided by a vendor which has ADM capabilities and may use these capabilities fully, partly or not at all.	

Extent of disclosure

Q.12	**When making disclosures about ADM, the OAIC's view is that entities must ensure that meaningful and accessible information is provided to consumers. In doing so, the OAIC considers that entities should strike an appropriate balance to ensure that disclosures are:** • **clearly articulated in plain language and easy to understand; structured to enable consumers to request further information, where required;** • **appropriately tailored – sufficiently specific to be meaningful, while avoiding overwhelming levels of detail;** • **organised so that similar information is grouped in a logical manner; and** • **framed in a way that allows the information and the decision to be challenged or contested.** **Having regard to the above considerations, is there anything additional that entities should disclose to provide effective and meaningful disclosures about ADM?**
In addition to the above, it would be valuable to outline: • the stages of a process that involve use of ADM • when and where in the process human centred oversight and decision making is utilised	

- what data and factors drive outcomes.

The inclusion of information regarding audit arrangements and mechanisms for review and challenge if any, to assess the performance of AI tools used for ADM should also be considered. This will ensure they are being used as intended and their outputs are consistent with initial scoping/testing. This is important to increase public trust and confidence.

Finally, there needs to be consideration as to the nature of the output and the value that transparency of this provides for consumers or the public. Where ADM is used within an existing automated process such as a document verification service or, in the case of digital health, automated checking of individual health identifiers in health provision, the output is often a simple determination as to whether information in two systems correctly match. In circumstances such as this, the output does not contribute to a risk assessment or value-based judgement of the consumer, and therefore disclosure of this use of ADM would likely be considered less meaningful than disclosure of the use of ADM to determine whether a person is eligible to be registered as a healthcare recipient in the My Health Record system.