Submission by Dr Rita Matulionyte

to the Department of Industry, Science and Resources

in response to Safe and Responsible AI Discussion Paper

25 July 2023

Dr Rita Matulionyte
Senior Lecturer, Macquarie Law School, Macquarie University
Room 512, 6 First Walk, Macquarie University, NSW 2109 Australia
Rita.matulionyte@mq.edu.au

About the author

I am a senior lecturer and researcher in technology law and intellectual property (IP) law at Macquarie Law School, Macquarie University. I am a recipient of Women in AI Award – Law Category (2023), an affiliate at the ARC Centre of Excellence for Automated Decision Making and Society, a lead of the Explainable AI Research Stream at the Centre for Applied Artificial Intelligence at Macquarie University, a Lead of the Emerging Technologies Workstream at the Australasian Society for Computers and Law, and an active member of the Australian Alliance for AI in Healthcare, Safety and Ethics Working Group.

I have been researching IP and technology law for the last 15 years, and have published over 50 peer-reviewed publications, presented numerous conference papers and prepared several reports for national and regional government bodies in this field. During the last 4 years, my research has focused on the **regulation and governance of AI technologies** in various sectors (government, healthcare, creative industries).

This submission is partly based on the following research papers:

1. R Matulionyte, 'Transparency of Facial Recognition Technology and Trade Secrets', in Rita Matulionyte and Monika Zalnieriute (eds), *Cambridge Handbook on Facial Recognition Technologies in the Modern State* (Cambridge University Press, forthcoming 2024)
2. R Matulionyte, T Abramovich, 'AI Explainability and Trade Secrets' in R Abbot (ed) *Research Handbook on Artificial Intelligence and Intellectual Property* 404-421 (Edward Elgar, 2022)
3. R Matulionyte, 'Increasing Transparency around Facial Recognition Technologies in Law Enforcement: Towards a Model Framework', *Information and Communication Technology Law* (forthcoming 2023), available on https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4408997
4. T Aranovich, R Matulionyte, 'Ensuring AI Explainability in Healthcare: Problems and Possible Policy Solutions', *Information and Communication Technology Law*, 2022, https://doi.org/10.1080/13600834.2022.2146395
5. R Matulionyte, P Nolan, F Magrabi, A Beheshti, 'Should AI Medical Devices be Explainable?', 30(2) *International Journal of Law and Information Technology*, 151-180 (2022)
6. R Matulionyte, 'Reconciling trade secrets and explainable AI: face recognition technology as a case study', 44(1) *European Intellectual Property Review 36* (2022)
7. R Matulionyte, 'Australian copyright law impedes the development of Artificial Intelligence: What are the options?', 52(4) *International Review for Intellectual Property and Competition Law -ICC* 417-443 (2021)
8. R Matulionyte, 'Can AI infringe moral rights of authors and should we do anything about it? An Australian perspective', *Law, Innovation and Technology* (forthcoming 2023), original manuscript available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4016001
9. R Matulionyte, Lee, J. 'Copyright in AI-generated works: lessons from recent developments in patent law' 19(1) *SCRIPTed – A Journal of Law, Technology and Society* 5 (2022)
10. N Selvadurai, R Matulionyte, 'Reconsidering Creativity: Copyright Protection for Works Generated Using Artificial Intelligence', *Journal of Intellectual Property Law and Practice* (2020) 15(7), 536
11. C White, R Matulionyte, 'Artificial Intelligence Painting a Larger Picture on Copyright,' *Australian Intellectual Property Review* (2020) 30, 224
12. Matulionyte, R., 'AI inventiveness and DABUS Saga', in Georg Nolte et all (eds), *Gestaltung der Informationsordnung. Festschrift fur Thomas Dreier zum 65. Geburtstag* 233-249 (Beck Verlag, 2022)
13. R Matulionyte, 'AI Inventor: Has the Federal Court of Australia Erred in its Decision in DABUS?', 13(2) *Journal of Intellectual Property, Information Technology and E-Commerce- JIPITEC* 1 (2022)

1. **Introduction**

1.1. I welcome the initiative taken by the Department of Industry, Science and Resources to seek input through *the Safe and Responsible AI Discussion Paper* 2023. A solid public consultation is clearly necessary given the societal implications of AI technologies and automated decision making (ADM) more generally.

1.2. These submissions are intended to be made public.

1.3. This submission focuses on selected **Questions 9 and 2** of the Issues Paper and specifically addresses two issues: **transparency of AI**, and **AI and copyright law** issues.

Consultation Question

9: Given the importance of transparency across the AI lifecycle, please share your thoughts on:

- ***a. where and when transparency will be most critical and valuable to mitigate potential AI risks and to improve public trust and confidence in AI?***
- ***b. mandating transparency requirements across the private and public sectors, including how these requirements could be implemented.***

Transparency around AI and automated decision making (ADM) is critical, both in private and public sectors. If properly implemented and enforced, transparency has a potential to improve quality and safe use of AI/ADM technologies and ensure accountability when AI is developed or used inappropriately.

Despite the recognition and establishment of the AI transparency principle in both the Australia's AI Ethics Framework and states' ethical AI policy documents (e.g. NSW AI Assurance Framework), Government itself acknowledges the continuing lack of transparency around AI technologies, including in government sector.[1] We thus welcome Government's interest in improving the implementation of this principle in practice.

It is well known that the need for transparency around AI/ADM systems will differ depending on the type of technology, where and how it is used, stakeholder needs in a specific context and the goals that government aims to achieve through specific

[1] E.g. NSW Ombudsman, *The New Machinery of Government: Using machine technology in administrative decision-making*, 29 November 2021, available at https://www.ombo.nsw.gov.au/__data/assets/pdf_file/0003/138207/The-new-machinery-of-government-special-report_Front-section.pdf, p 7

transparency measures.[2] There is therefore no single or easy answer as to where and how AI transparency principle should be implemented.

As a starting point, we propose the following guiding principles when establishing transparency duties with relation to AI/ADMS in specific contexts:

1. Transparency is important when AI/ADM systems are used *both* in **public** and **private** sectors.
2. More transparency is needed when the risk/impact of the AI/ADM system is higher, and less transparency is needed when it is lower. Such **risk-based approach** to AI transparency has been adopted in the proposed EU AI Act, Canada's Directive on Automated Decision Making and elsewhere.
    a. Note: while transparency will thus be most critical and valuable in high-risk scenarios (e.g. AI in healthcare or government), transparency alone is *not* sufficient to ensure quality and safe use of such technologies and thus should be only one integral part of a comprehensive quality assurance framework.
3. Transparency around AI/ADMS will have to be implemented both through **sector-specific and horizontal AI legislation**:
    a. *Sector-specific legislation*: e.g. the EU Digital Services Act sets specific transparency duties for online service providers or online services.[3] Similarly, in Australia, transparency duties will have to be implemented in sector specific legislation (e.g. privacy, financial services, medical device legislation and others).
    b. *Horizontal legislation on AI*: e.g. in the draft EU AI Act horizontal transparency duties are set to all AI systems, but they vary depending on the level of risk of the AI system. If Australia were to adopt a horizontal regulation of AI, it would have to include base-line transparency standards.
4. Transparency of AI in **private and public sectors** could be achieved by revising **existing instruments** and **introducing new measures**.

    ***Government/public sector*:**

    a. transparency around AI/ADMS could be improved by revising freedom of information legislation (federal and state level). Such legislation could clearly establish what information government agencies should disclose about ADMS they use and how the disclosure should be made (e.g. through proactive publication of information online or upon request from stakeholders). These laws, together with improved government procurement practices, should also aim to address barriers

[2] AI transparency and explainability in different sectors is discussed in e.g. R Matulionyte, P Nolan, F Magrabi, A Beheshti, 'Should AI Medical Devices be Explainable?', 30(2) *International Journal of Law and Information Technology*, 151-180 (2022); R Matulionyte, 'Increasing Transparency around Facial Recognition Technologies in Law Enforcement: Towards a Model Framework' (forthcoming in Information and Communications Technology Law in 2024). Available at SSRN: https://ssrn.com/abstract=4408997 or http://dx.doi.org/10.2139/ssrn.4408997

[3] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act), *OJ L 277, 27.10.2022, p. 1–102*, Arts 15, 27, 39.

towards ensuring government transparency, such as trade secrets (see below).[4]

b. The government should consider creating a register of AI/ADMS systems, at least for high-risk systems and/or those used by government agencies. International experiences, such as the UK Algorithmic Transparency Recording Standard,[5] and the proposed EU Register for high-risk AI[6] could be used as examples.
c. Specific transparency rules will need to be designed for specific government sectors, such as law enforcement or judiciary (administration of justice) where the use of AI might lead to especially significant impacts on human rights. For example, Interpol/ WEF Guidelines provide detailed transparency requirements for the use of facial recognition technologies in law enforcement context.[7]

***Private sector**:*

d. Stronger transparency duties around ADMS need to be discussed in domain-specific legislation, especially in sectors where AI may pose significant risks to health, safety or human rights (e.g. medical device, financial services, transport regulation, etc).
e. Horizonal law on AI, if considered for Australia, could establish more general risk-dependent transparency duties that would apply to private stakeholders not covered by sector-specific legislation (see horizontal transparency duties under the proposed EU AI Act).

When discussing how transparency around AI/ADMS could be implemented in various instruments and initiatives, the following **5 steps** are recommended:

Step 1: Agreeing on the definition of 'AI transparency' and related terms

There is currently a lack of consensus, both in Australia and internationally, what 'AI transparency' concept means and how it relates to other concepts, especially AI explainability. For instance, the Australia's AI Ethics Framework[8] and the Digital Transformation Agency's (DTA) Guidance[9] refers to 'transparency and explainability'; the NSW AI Assurance Framework mentions 'transparency' only (and closely ties it with contestability);[10] while the Commonwealth Ombudsman requires

[4] See R Matulionyte, 'Government Automation, Transparency and Trade Secrets', under review, request a copy from author: rita.matulionyte@mq.edu.au
[5] Available at https://www.gov.uk/government/collections/algorithmic-transparency-recording-standard-hub.
[6] See draft EU AI Act.
[7] World Economic Forum et al, *Policy Framework for Responsible Limits of Facial Recognition Technology. Case study: law enforcement*, Revised 2022, https://www.weforum.org/reports/a-policy-framework-for-responsible-limits-on-facial-recognition-use-case-law-enforcement-investigations-revised-2022.
[8] Australia's Artificial Intelligence Ethics Framework in 2019.
[9] Digital Transformation Agency (DTA), Guidance for Adoption of AI in the Public Sector https://www.architecture.dta.gov.au/sp_aga2?id=aga_capability_display&kb=KB0011078.
[10] NSW Government's AI Assurance Framework

AI to be 'understandable'. In order to have a constructive policy discussion, the Government should agree on the common terminology and what meaning they attribute to each term they use in the discussion.

For instance, the Government could use the definitions of 'transparency' and 'explainability' provided in the IEEE Standard for Transparency of Autonomous Systems 7001-2021, which is one of the most recent international documents on the topic and which is being adopted in most recent policy/legislative solutions in national jurisdictions (e.g. UK).[11] The IEEE Standard 7001-2021 provides the following definitions:

transparency: A transfer of information from an autonomous system or its designers to a stakeholder that is truthful; contains information relevant to the causes of some action, decision, or behavior; and is presented at a level of abstraction and in a form meaningful to the stakeholder. Transparency should be mindful of the stakeholders' likely perception and comprehension, and should avoid disclosing information in a manner that, while technically true, is framed in a way that leads to misapprehension.
explainability: The extent to which the information made transparently available to a stakeholder can be readily interpreted and understood by a stakeholder.

Whichever terminology or definition is chosen, it is important that it is used *consistently* across federal and state governments to avoid unnecessary confusion and ensure a constructive debate.

Step 2: Setting Appropriate and Achievable Goals for AI/ADMS Transparency

Before deciding where, when and how to ensure AI transparency, it is necessary to identify the goals that transparency measures are supposed to achieve.

There should be some evidence that the identified goals could be achieved with the help of transparency, that there is no other more efficient way in achieving these goals and, preferably, they should be measurable. For instance, research suggests that in some cases, model-level transparency is preferable to technical explainability of AI outputs as technical explainability might be little helpful in achieving trust or quality assurance.[12]

Transparency is not an end in itself. Transparency can serve many purposes, such as increasing trust in AI, ensuring contestability of decisions produced with the help of AI, enabling external scrutiny and quality assurance of AI, and accountability by responsible stakeholders (developers, users of AI). Depending on the goal chosen, different transparency measures will be needed.

For instance, if the goal of specific transparency rules is to increase trust in AI among a general non-expert public, then general information about the purpose of

---

https://www.digital.nsw.gov.au/policy/artificial-intelligence/nsw-artificial-intelligence-assurance-framework.

[11] E.g. definition adopted in this standard was implemented in the UK Algorithmic Transparency Recording Standard, version 2.1
available at https://www.gov.uk/government/publications/algorithmic-transparency-template

[12] see R Matulionyte, P Nolan, F Magrabi, A Beheshti, 'Should AI Medical Devices be Explainable?', 30(2) *International Journal of Law and Information Technology*, 151-180 (2022).

the AI tool, how its risks have been addressed would be sufficient. If the goal of transparency is to ensure public scrutiny that could lead to accountability for inappropriate use of AI, then stakeholders should be required to provide more detailed technical and legal information about a specific AI tool that would enable external stakeholders (expert members of the public, e.g. NGOs and university researchers) to identify possible technical and/or legal defects of the technology, flag them and monitor that they are addressed and, where applicable, ensure that responsible private or public actors are held accountable if the technology is developed or used inappropriately.

While different goals – trust, individual contestability, quality assurance, accountability, etc. – are legitimate and reasonable ones, Government could consider prioritizing some of them, which would make it easier to design better **targeted** transparency policies and measure their effectiveness. For instance, the UK Algorithmic Transparency Recording Standard lists public scrutiny, accountability and trust as the three primary goals of algorithmic transparency.[13]

It is recommended to avoid vague goals – such as 'mitigating the risks of AI' (as mentioned in the Discussion Paper), and identify more **specific goals**, such as enabling public scrutiny of AI technologies, facilitating accountability for inappropriate use of AI tools etc.

Step 3: Identifying Stakeholders Who Need Information About AI and Why

Third, depending on the goals of transparency measures, Government needs to clearly identify which stakeholders the transparency measures will target. Various stakeholders are generally interested in transparency, e.g.:[14]

- General public, including both non-experts and expert public members
- Individuals directly affected by an AI output
- Courts and litigants during litigation regarding AI related harms
- Regulatory, certification, auditing authorities
- Independent experts, including community organizations and university-based researchers.

Different stakeholders will need different types of information/ levels of transparency that would correspond to their AI understanding levels and different goals they seek to achieve. It is important to ensure that legitimate information/transparency needs of all relevant stakeholders could be satisfied, i.e. information suitable and relevant for various stakeholders is produced/documented and is made available either proactively or upon request (reactively).

At the same time, it would be unreasonable to expect that *all* stakeholders are *always* provided relevant information about *all* AI technologies, especially when AI technology at stake is a low-risk one. This would lead to **information overload** and may lead to 'transparency fallacy'. Too extensive transparency duties would also

[13] Available at https://www.gov.uk/government/publications/algorithmic-transparency-template te
[14] Similar: Australian Artificial intelligence Ethics Framework lists various stakeholders interested in transparency.

lead to **high costs** for AI developers and deployers which would be translated into higher prices charged from end users/consumers and/or to slower adoption of AI.

It is therefore suggested that, when setting transparency standards through specific legal or governance frameworks, the Government should carefully assess which stakeholders' transparency needs should be met as a priority, i.e. who will be the **primary beneficiaries** of the transparency measures.

Step 4: Defining the Nature and Scope of Information to be Provided

The next essential step is to define what information needs to be provided about specific technologies and how.

In terms of the types of information that various stakeholders might need, it could be, among other things:

- Information that individuals are interacting with an AI system;
- Information that a decision was made with the assistance of an AI system; what role AI played and what parameters it took into account when making the decision;
- Details on the capabilities, limitations, safety and risks of a specific AI system;
- Details on the training and validation of the AI system, e.g. methodology used to train the AI systems; datasets used in pre-training, fine-tuning, etc; validation and real-life testing information.

Regulatory or governance instruments could set different pathways for providing information, e.g.

- Proactively – when AI developer or deployer is required to proactively publish certain information about AI (on websites, registers, in labels, user manuals, decisions issued to individuals)
- Reactively – when information is required to be disclosed upon a request from an interested stakeholder.

We recommend that the scope of transparency required should differ depending on the **risk/impact level of the technology**, i.e. the higher risk/impact of technology, the more transparency is required.

As noted above, this **risk approach** has been adopted in e.g. the draft EU AI Act. It requires minimum information about low-risk AI systems (letting users know about the system) and sets high transparency duties for high-risk systems (e.g. registering in the EU AI Register).[15] The information provided should be related to the general and specific risks of the system, i.e. enable stakeholders to assess whether risks have been properly identified and mitigated.

Step 5: Address Barriers to Effective Transparency

[15] See also UK Algorithmic Transparency Recording Standard which sets transparency requirements only with relation to systems that meet certain risk/impact threshold.

In some situations, information desired by stakeholders cannot be accessible or disclosed due to privacy, confidentiality, security and other reasons.[16] Commercial confidential information (trade secrets) is becoming an especially important barrier in ensuring transparency around AI.

In *O'Brien* decision,[17] trade secrets were one of the main reasons why an individual was rejected access to information about the algorithm that denied them social housing benefits.

Government needs to identify and **carefully assess barriers** in ensuring effective transparency, including trade secrets, and how they could be removed or mitigated.

While trade secret protection cannot and should not be entirely removed, there are ways how exceeding trade secret claims by AI developers and vendors could be addressed. For instance, laws or procurement contracts could require certain *essential* algorithmic information to be disclosed to certain stakeholders (e.g. regulatory authorities, independent experts) or to the entire public, despite its possible confidential status.[18]

Consultation Question

2. What potential risks from AI are not covered by Australia's existing regulatory approaches? Do you have suggestions for possible regulatory action to mitigate these risks?

One of the risks that current Australian laws do not properly address is **copyright infringement.** This challenge has been in-more detail discussed in another policy submission,[19] which is briefly summarized below.

One the one hand, the unauthorized use of copyright-protected content (text, images, videos, etc) in ML training process may lead to copyright infringement and liability of AI developers. [20]

On the other hand, even if AI developers were willing to license and pay for the use of such content in ML training, there are frequently no viable licensing schemes that

[16] See R Matulionyte, T Abramovich, 'AI Explainability and Trade Secrets' in R Abbot (ed) *Research Handbook on Artificial Intelligence and Intellectual Property* 404-421 (Edward Elgar, 2022); R Matulionyte, 'Reconciling trade secrets and explainable AI: face recognition technology as a case study', 44(1) *European Intellectual Property Review 36* (2022); 1.
[17] *O'Brien v Secretary, Department Communities and Justice* [2022] NSWCATAD 100.
[18] See R Matulionyte, 'Government Automation, Transparency and Trade Secrets', note 4.
[19] See R Matulionyte, Submission to Copyright Enforcement Review (2023), available at https://consultations.ag.gov.au/rights-and-protections/copyright-enforcement-review/consultation/view_respondent?uuId=192576760
[20] See R Matulionyte, 'Australian copyright law impedes the development of Artificial Intelligence: What are the options?', 52(4) *International Review for Intellectual Property and Competition Law -ICC* 417-443 (2021).

would allow AI developers to acquire copyright licenses to millions or billions of content pieces (works) needed to train ML modules.

In comparison, copyright laws in other comparable jurisdictions have **copyright exceptions**, such as fair use (US) or 'text and data mining' exception (UK, EU) that allow free use of content for AI training purposes at least in certain contexts (e.g. for non-commercial purposes). These exceptions are currently being tested in courts.[21]

The Attorney-General's Department is currently consulting with stakeholders on whether and how Australian copyright law should be revised in light of challenges raised by AI technologies. We encourage to continue the discussion and **establish solutions** which both promote legal certainty for local AI industries and consider the legitimate interests of copyright holders.

The main **recommendations** in the field are:

1. Discuss and develop measures to **improve transparency** around the use of copyright-protected content in ML context, which would enable right holders to identify when their works are used in ML process.
2. Examine the **effectiveness of** current individual and collective copyright **licensing options** for the use of copyright-protected content in ML context, and collect best international practices in the field, with the goal of improving licensing mechanisms in Australia.
3. Continue a policy debate on what **copyright exception** would be most suitable to balance both the interests of AI developers and right holders, which would both encourage innovation in AI space and ensure remuneration for right holders, at least in case of commercial AI applications.

[21] Currently, there are at least 4 lawsuits pending against AI developers who used copyright-protected content to train their AI modules. For an overview see R Matulionyte, Submission to Copyright Enforcement Review (2023), available at https://consultations.ag.gov.au/rights-and-protections/copyright-enforcement-review/consultation/view_respondent?uuId=192576760