



## King's Research Portal

DOI:

<https://doi.org/10.18742/pub01-104>

[Link to publication record in King's Research Portal](#)

*Citation for published version (APA):*

(Jonak) Lisinska, J. (Ed.) (2022, Nov 25). Response to Parliamentary inquiry on Governance of artificial intelligence (AI). <https://doi.org/10.18742/pub01-104>

### **Citing this paper**

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

### **General rights**

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

### **Take down policy**

If you believe that this document breaches copyright please contact [librarypure@kcl.ac.uk](mailto:librarypure@kcl.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

# Response to [Parliamentary inquiry](#) on Governance of artificial intelligence (AI)

## Introduction about TAS-hub/Governance Node and Functionally Node

### **About TAS-hub**

The UKRI TAS Hub (EP/V00784X/1), assembles a team from the Universities of Southampton, Nottingham and King's College London. The Hub sits at the centre of the £33M Trustworthy Autonomous Systems Programme, funded by the UKRI Strategic Priorities Fund. The role of the TAS Hub is to coordinate and work with six research nodes to establish a collaborative platform for the UK to enable the development of socially beneficial autonomous systems that are both trustworthy in principle and trusted in practice.

### **About Governance Node**

The UKRI TAS Node in Governance & Regulation (EP/V026607/1) is a multidisciplinary team working on modes of formal and informal governance on autonomous systems. It explores notions of accountability and responsibility and works on designing tools enabling developers and regulators to embed these values in AI systems.

### **About Functionality Node**

The UKRI TAS Node in Functionality (EP/V026518/1) is a research programme that aims to create 'Design-for-Trustworthiness' methods and guidelines for autonomous systems that have the ability to adapt and evolve in functionality. The functionality node has four work packages (Specification, Ethics, Regulation and Verification), focusing on how to specify, verify, regulate and consider the ethical aspects of three other work packages that are involved in developing technologies with different methods of adaptation (Swarm Robotics, Soft Robotics, and Uncrewed Air Vehicles).

Citation: TAS-Hub, Governance Node and Functionality Node (2022) Response to Parliamentary Inquiry on Governance of artificial intelligence (AI). DOI: 10.18742/pub01-104

## Authors

Downer John, Krook Joshua, Lindley Joseph, Lisinska Justyna, Moreau Luc, Onitiu Daria, Ramchurn Gopal, Simperl Elena, Walker Johanna, Winter Peter

## Executive summary and our recommendations:

### **Effectiveness of the current governance of AI in the UK**

- Defining standards for clearly and explicitly referring to how specific instances of AI work and what they do, including distinguishing between models, data and algorithms, will enhance effective governance.
- Promoting the wider adoption of Responsible AI Licenses (RAILs) by mandating their use in public sector AI applications can support a pro-innovation governance approach.

### **Measures to make AI more explainable and transparent to the public**

- Further clarification regarding the role of the cross-sectoral principles and context-driven approach in the AI policy proposal is needed to enable coherence across cross-regulatory landscapes and support for regulatory bodies. In particular, the notions of transparency and explainability need to be mandatory and address potential impact to ensure accountability.
- Citizen engagement and awareness raising around AI is key for both use and as protection against misuse of AI; in particular, approaches that both educate and entertain through the arts and culture can be used. This approach can reach a wide range of citizens. The use of AI in the arts to counteract online misinformation and disinformation has already been established and has a more inclusive reach.<sup>1</sup>

### **Legal framework for the use of AI and its fitness for purpose**

- More support needs to be provided for regulators to assist in producing domain-specific legal frameworks in accordance with the 2022 principles while ensuring their frameworks are fit for purpose and fair to the humans to whom AI-related liabilities might accrue.
- Regulators would benefit from specific legal frameworks to assign liability, as we saw in healthcare and autonomous vehicles domains.

### **Lessons that the UK government can learn from the Australian context**

- Australia has taken a principles-based, low-regulatory approach to artificial intelligence, but a lack of laws on AI has led to widespread mistrust of its use by government and private companies.
- The current government is exploring new laws on AI and reviewing existing privacy legislation.

---

<sup>1</sup> [www.mediafutures.eu](http://www.mediafutures.eu)

- Government agencies have been found to be in breach of the government's own AI Ethics Principles. This shows why a hands-off, non-regulatory approach to AI may be insufficient in ensuring compliance.
- The Australian context demonstrates that AI regulation should be tied into broader laws around human rights, civic rights and individual freedoms to allow for judicial review of automated decisions.

### **Lessons from the EU context**

- The EU defines AI separately from the technologies that constitute it, leaving flexibility for growth and development;
- The EU does not have a 'one size fits all' approach to AI regulation but rather has four categories of AI Risk and approaches each of these differently;
- There are multiple governance tools for organisations to confirm with the EU AI Act in existence.

## Responses

### **How effective is the current governance of AI in the UK?**

Contact details:

*Joseph Lindley, Senior Research Fellow, Security Lancaster, University of Lancaster.*

[j.lindley@lancaster.ac.uk](mailto:j.lindley@lancaster.ac.uk)

*Johanna Walker, Research Associate, Distributed Artificial Intelligence Group, Department of Informatics, King's College London, johanna.walker@kcl.ac.uk*

We recommend the use of standards to support clear, unambiguous descriptions of how AI is deployed and the promotion of Responsible AI Licencing to support pro-innovation governance of AI.

The term AI is inherently ambiguous and can be misleading, posing problems for governance. In a report commissioned by the Arts and Humanities Research Council<sup>2</sup> domain experts noted:

- "It is sexy to put the term on stuff" (Anab Jain, Superflux)
- "[AI can be] just processing without any kind of actual learning" (Professor Ann Light, University of Sussex)
- "Don't use the terminology [...] it's so vague" (Professor Nicholas Nova, Near Future Laboratory)
- "Any in-depth conversation about AI will result in the dialogue becoming about sentient killer robots" (Dr Joseph Lindley, Lancaster University)

Because anyone can term anything to be AI, products and services described as AI could:

---

<sup>2</sup> Lindley, J., & Coulton, P. (2020). AHRC Challenges of the Future: AI & Data Report.

- Potentially have no ‘intelligent’ component whatsoever.
- Use a static AI that leverages probability to make predictions.
- Use advanced machine learning techniques to find solutions to problems that are unexplainable to humans.

The July 2022 policy paper *Establishing a pro-innovation approach to regulating AI*<sup>3</sup> acknowledges the challenge of meaningful definition in the context of governance and proposes using two characteristics:

- adaptiveness (the ability to behave in ways that were not explicitly programmed);
- and autonomy (the ability to act without human intervention).

The approach has virtues, but an obvious shortcoming is that many products and services that do not require any special considerations regarding AI governance would still be included. For example, something as simple as a radio alarm clock that plays an automatically-generated list of suggested songs is both adaptive and autonomous.

The value of standardisation and supported by unambiguous explanations of what a particular technology or system does is demonstrated in DCMS commissioned report relating to internet-connected devices<sup>4</sup>. The use of simple labelling systems may significantly enhance the legibility of AI systems by making how they function visible, without having to directly engage with the challenge of defining AI<sup>5</sup>. Such approaches can underpin trusting relationships with AI-infused products and services while minimising the need for interventionist regulatory approaches. However, the EU Artificial Intelligence Act, which is applicable to Northern Ireland, states that “the notion of AI system should be clearly defined to ensure legal certainty while providing the flexibility to accommodate future technological developments”.<sup>6</sup>

Responsible AI Licences (‘RAILs’, see<sup>7</sup>) utilise the well-established mechanisms of software and technology licensing to promote self-governance within the AI sector. RAILs allow developers, researchers, and companies to publish AI innovations while specifying restrictions on the use of source code, data, and models. These restrictions can refer to high-level restrictions (e.g., prohibiting uses that would discriminate against any individual) as well as application-specific restrictions (e.g., prohibiting the use of a facial recognition system without consent). RAILs have recently been adopted in BLOOM (the world’s largest AI language model) and the Stable Diffusion image-generating AI. The adoption of such licenses for AI systems funded by public

---

<sup>3</sup> Office for Artificial Intelligence. Establishing a pro-innovation approach to regulating AI. <https://www.gov.uk/government/publications/establishing-a-pro-innovation-approach-to-regulating-ai/establishing-a-pro-innovation-approach-to-regulating-ai-policy-statement>

<sup>4</sup> Office for Artificial Intelligence. Establishing a pro-innovation approach to regulating AI. <https://www.gov.uk/government/publications/establishing-a-pro-innovation-approach-to-regulating-ai/establishing-a-pro-innovation-approach-to-regulating-ai-policy-statement>

<sup>5</sup> Lindley, J., Akmal, H. A., Pilling, F., & Coulton, P. (2020). Researching AI Legibility through Design. Proceedings of the 2020 CHI Conference on Human Factors in Computer Systems - CHI '20. <https://doi.org/10.1145/3313831.3376792>

<sup>6</sup> <https://artificialintelligenceact.eu/the-act/>

<sup>7</sup> Ferrandis, C. M., & Contractor, D. (2022). BigScience Open Rail-M License. <https://www.licenses.ai/blog/2022/8/26/bigscience-open-rail-m-license>

procurement and publicly-funded AI research will help support a pro-innovation culture that acknowledges the unique governance challenges posed by emerging AI technologies.

Given the difficulties of identifying and creating boundaries around a single term, it may be useful to consider the component elements separately - the data used, the model or algorithm, and the implementation. Each area carries risk, and examining 'AI' as an amorphous concept creates a potential for overlooking exactly what governance needs to address. Implementation risks include potential 'dual use' risks (which are already applicable in Northern Ireland)<sup>8</sup>.

### **What measures could make the use of AI more transparent and explainable to the public?**

Contact details:

*Daria Onitiu, Research Associate, Governance & Regulation Node, Edinburgh Law School, [donitiu@ed.ac.uk](mailto:donitiu@ed.ac.uk)*

*Johanna Walker, Research Associate, Distributed Artificial Intelligence Group, Department of Informatics, King's College London, [johanna.walker@kcl.ac.uk](mailto:johanna.walker@kcl.ac.uk)*

The main concern with the AI policy innovation paper includes the interplay of the context-driven approach with the cross-sectoral principles in ensuring that notions of transparency and explainability are fit for purpose across cross-regulatory bodies. We welcome recent guidance by the Cabinet Office's Central Digital and Data Office, the Information Commissioner's Office, and the Financial Conduct Authority with the Alan Turing Institute who strengthen the context-driven as well as a horizontal approach to AI governance and highlighting the role of responsible innovation for AI design and deployment<sup>9</sup>. We also note already some divergence between the AI policy document's approach and the guidance mentioned:

- ***The emphasis on potential impacts:*** contrary to the AI policy proposal, which emphasises a pro-innovation and risk-based approach focusing on high-risk, current guidance by regulatory guidance emphasises the role of transparency and

---

<sup>8</sup> Department for International Trade (2021) <https://www.gov.uk/government/publications/notice-to-exporters-202112-new-dual-use-regulation-eu-2021821/note-202112-new-dual-use-regulation-eu-2021821>

<sup>9</sup> ICO (2022) What are the accountability and governance implications of AI?. <https://ico.org.uk/for-organisations/guide-to-data-protection/key-dp-themes/guidance-on-ai-and-data-protection/what-are-the-accountability-and-governance-implications-of-ai/>

Ostmann, Florian; Dorobantu, Cosmina. (2021). AI in Financial Services. Zenodo. <https://doi.org/10.5281/zenodo.4916041>

Cabinet Office Central Digital and Data Office (2021). Algorithmic transparency data standard. [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/1037941/Algorithmic\\_transparency\\_data\\_standard.csv/preview](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1037941/Algorithmic_transparency_data_standard.csv/preview)

Aitken, M., Leslie, D., Ostmann, F., Pratt, J., Margetts, H., & Dorobantu, C. (2022). Common Regulatory Capacity for AI. The Alan Turing Institute. <https://doi.org/10.5281/zenodo.6838946>

explainability to mitigate potential impacts<sup>10</sup>. Our view resonates with measures for AI governance, obliging manufacturers, developers, and providers to consider emergent risks and potential uses and misuses of emerging technology. To avoid any discrepancy between the role of cross-sectoral principles as a standard-setting and a risk-based tool, there needs to be an update of the language and normative goals of transparency and explainability for both high-stake decisions and the type of algorithms employing self-learning capabilities.

- ***The role of mandatory transparency obligations and “prohibited” practices:*** We welcome the AI policy document’s statement whereby ‘regulators may deem that decisions which cannot be explained should be prohibited entirely’. This implies that transparency obligations will become mandatory obligations for certain “high-risk decisions” and/or policies that will have a significant impact on people’s human rights. Several efforts - from guidance on challenges of transparency (i.e. model cards), specific use-cases on explainable AI, to sector-specific projects on providing interpretability in high-stake decisions (such as, [MHRA “project glass-box”](#)) - intend to cover different facets of the notions of transparency and accountability, as well as explainability<sup>11</sup>

Having said that, notable differences will emerge in how (mandatory) transparency obligations will be enforced in practice. For example, we see a significant need for capacity-building for public authorities who will need to demonstrate accountability, as well as notions of explainability and fairness. The Central Digital and Data Office and Office for Artificial Intelligence only stipulate requirements to industry (i.e. developers and data scientists) that are process-based<sup>12</sup>. Further, the Centre for Data Ethics and Innovation recognises that public sector organisations might need further guidance on the interplay between data protection, non-discrimination, as well as sectoral rules to mitigate algorithmic bias and ensure algorithmic fairness, considering the public sector equality duty in the Equality Act 2010<sup>13</sup>. Therefore, we suggest more guidance enabling regulatory bodies’ capacity building to oversee process-based standards and more substantive guidance in how standard-setting fits with current legislative frameworks.

---

<sup>10</sup> Ostmann, Florian; Dorobantu, Cosmina. (2021). AI in Financial Services. Zenodo.

<https://doi.org/10.5281/zenodo.4916041>

Cabinet Office Central Digital and Data Office (2021). Algorithmic transparency data standard. [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/1037941/Algorithmic\\_transparency\\_data\\_standard.csv/preview](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1037941/Algorithmic_transparency_data_standard.csv/preview)

<sup>11</sup> <https://digital-transformation.hee.nhs.uk/building-a-digital-workforce/dart-ed/horizon-scanning/understanding-healthcare-workers-confidence-in-ai/chapter-3-governance/guidelines>

NHS AI Lab & Health Education England (2022). Understanding healthcare workers’ confidence in AI.

<https://digital-transformation.hee.nhs.uk/binaries/content/assets/digital-transformation/dart-ed/understandingconfidenceinai-may22.pdf>

<https://modelcards.withgoogle.com/about>

Leslie et al (2021). Explaining decisions made with ai: a workbook.

<https://arxiv.org/ftp/arxiv/papers/2104/2104.03906.pdf>

MHRA (2022). Software and AI as a Medical Device Change Programme - Roadmap.

<https://www.gov.uk/government/publications/software-and-ai-as-a-medical-device-change-programme/software-and-ai-as-a-medical-device-change-programme-roadmap>

<sup>12</sup> Central Digital and Data Office and Office for Artificial Intelligence (2019) Understanding artificial intelligence ethics and safety. <https://www.gov.uk/guidance/understanding-artificial-intelligence-ethics-and-safety>

<sup>13</sup> Centre for Data Ethics and Innovation (2020). Review into bias in algorithmic decision-making.

<https://www.gov.uk/government/publications/cdei-publishes-review-into-bias-in-algorithmic-decision-making/main-report-cdei-review-into-bias-in-algorithmic-decision-making#transparency-in-the-public-sector>

## Improving citizen awareness and data/AI literacy

Demand-side investment is necessary as well as supply-side - improving data literacy in public in general and specifically as a soft skill for employment. Outside of those with a science or computing background, most people do not recognise what data is, so they are not aware of what input AI is using or how this might affect them<sup>14</sup>.

Awareness and literacy raising can be carried out effectively through the arts and media sectors. A fragmented, polarised public engagement and shortening attention spans challenge us to create new approaches, experiences and information technologies that entertain and educate at the same time. Citizen engagement and awareness raising around AI is key for both use and as protection against misuse of AI; in particular, approaches that both educate and entertain through the arts and culture can be used. This approach can reach a wide range of citizens. The use of AI in the arts to counteract online misinformation and disinformation has already been established, for instance, in artworks that educate audiences about deepfakes<sup>15</sup> or augmented reality dinosaurs to teach children about media literacy through gamification<sup>16</sup>.

## To what extent is the legal framework for the use of AI, especially in making decisions, fit for purpose?

Contact details:

*John Downer: Co-I Functionality Node and Senior Lecturer in Risk and Resilience, School of Sociology, Politics and International Studies (SPAIS), University of Bristol.*  
[john.downer@bristol.ac.uk](mailto:john.downer@bristol.ac.uk)

*Peter Winter: Research Associate, Functionality Node, School of Sociology, Politics and International Studies (SPAIS), University of Bristol.* [peter.winter@bristol.ac.uk](mailto:peter.winter@bristol.ac.uk)

The UK Government's publication of the *AI Regulation Policy Paper*<sup>17</sup> provides a general outline of the legal arrangements relating to the use of AI in the UK. It must be emphasised that the policy paper only outlines the legal foundations of the new structure, assessing that: '*accountability for the outcomes produced by AI systems and legal liability must always rest with an identified or identifiable legal person*'. While the 2022 paper does clarify liability by assigning it to the *human* responsible for *using* the AI system, it lacks clear guidelines on how each regulator will enforce this standard in practice. It could be argued that this ambiguity provides both stability *and* flexibility of regulation: allowing

---

<sup>14</sup> Johanna Walker et al (2021) Data Literacy: From Higher Education to the Workplace  
<https://dedalus.pa.itd.cnr.it/en/news/8-how-do-universities-and-companies-deal-with-data-literacy-report.html>

<sup>15</sup> <https://aniacatherine.com/soft-evidence>

<sup>16</sup> <https://mediafutures.eu/2nd-cohort-projects/ochi-media-literacy-edutainment-app/>

<sup>17</sup> United Kingdom Secretary of State for Digital, Culture, Media and Sport (2022) AI Regulation Policy Paper: Establishing a pro-innovation approach to regulating AI.  
[.https://www.gov.uk/government/publications/establishing-a-pro-innovation-approach-to-regulating-ai/establishing-a-pro-innovation-approach-to-regulating-ai-policy-statement](https://www.gov.uk/government/publications/establishing-a-pro-innovation-approach-to-regulating-ai/establishing-a-pro-innovation-approach-to-regulating-ai-policy-statement)



regulators to create domain-specific rules for their interpretation of the standard in light of contextual and situated activities specific to their organisation. At the same time, however, it potentially creates a challenge for some regulators who may lack the expertise to apply this principle in their sector or domain.

For this matter, we recommend that regulators enlist legal counsel to assist in producing domain-specific legal frameworks in accordance with the 2022 principles while ensuring their frameworks are fit for purpose and fair to the humans to whom AI-related liabilities might accrue. Legal collaboration should focus on the production of empirical research that ranges from helping regulators and their stakeholders judge whether the AI is safe (or unsafe) to how people are managed or organised within decision-making procedures. For example, the UK DoT's Centre for Connected and Autonomous Vehicles (CCAV) enlisted the Law Commission in the process of creating a legal framework for its Automated Vehicles project (note the term 'automated' as opposed to 'autonomous')<sup>18</sup>. Elsewhere, in the NHS, the 'NHS AI Lab'<sup>19</sup> has been created to establish a legal framework for various parties involved in developing, deploying and using AI systems in healthcare through its 'Liability and Accountability programme', drawing on both the expertise of NHS's expert legal panel and a collaborative programme set up to address a barrier to development and implementation.

There are many liabilities and related issues that may arise as an effect of AI-assisted decision-making. They give rise to complex questions regarding, for instance, the degree of real agency that individuals held responsible for AIs have over the outputs and defects of those AIs. Insofar as legal liability always rests with an identified or identifiable person, moreover, then there is good reason to expect pushback from humans using AI systems involved in collaborative decisions.

Research on how new technologies are governed or implemented in the workplace has a long history of highlighting pushback by workers, public(s) and activists<sup>20</sup>, with recent examples ranging from medical experts with concerns over the efficacy and safety of AI systems<sup>21</sup>, to customers of financial service companies with concerns regarding the competency of AI-driven chatbots<sup>22</sup>. Additionally, issues that relate to differing measures of agency also draw attention to the problem of how to ensure effective redress for victims. For example, in addition to forging legal frameworks, legal experts must work with

---

<sup>18</sup> Law Commission (2022) Automated Vehicles: Joint Report. <https://s3-eu-west-2.amazonaws.com/lawcom-prod-storage-11jsxou24uy7q/uploads/2022/01/Automated-vehicles-joint-report-cvr-03-02-22.pdf>

<sup>19</sup> NHS (2022) Understanding Healthcare Workers' Confidence in AI. <https://digital-transformation.hee.nhs.uk/building-a-digital-workforce/dart-ed/horizon-scanning/understanding-healthcare-workers-confidence-in-ai/chapter-3-governance/guidelines>

<sup>20</sup> Juma, C (2016) *Innovation and its Enemies: Why People Resist New Technologies*. New York, NY: Oxford University Press.

<sup>21</sup> Boggs, A. M (2019) Harnessing crash and disengagement data to analyse the safety performances of automated vehicles. PhD diss. University of Tennessee, 2019.

Oakes, K. (2020) "Radiologists to FDA: Autonomous AI not ready for prime time". *Regulatory Affairs Professionals Society*. <https://www.raps.org/news-and-articles/news-articles/2020/7/radiologists-to-fda-autonomous-ai-not-ready-for-pr>

<sup>22</sup> Luo, X., Tong, S., Fang, Z., Qu, Z (2019) Frontiers: Machines vs. Humans: The Impact of Artificial Intelligence Chatbot Disclosure on Customer Purchases. *Marketing Science* 38(6):937-947.  
Whittaker, T (2022) 'Artificial Intelligence Liability: Proposed EU Directive. <https://blog.burges-salmon.com/post/102hy48/artificial-intelligence-liability-proposed-eu-directive>

regulators and stakeholders (and the UK Government) to put in place measures to provide compensation in respect to damages. Regulators, organisations and legal experts must play a significant role in compensating victims for harms caused by AI systems, in setting standards, or in allocating blame. In spheres such as the European Commission, the Artificial Intelligence Liability Directive (AILD) has been promoted under a variety of Governments to ensure that "*persons harmed by artificial intelligence systems enjoy the same level of protection as persons harmed by other technologies*"<sup>23</sup>. However, despite the optimism of this EC announcement, legislative provisions to smooth the path to compensation for those harmed by AI systems are still under development across most domains<sup>24</sup>. The automotive domain, however, might be seen as an exception to this rule, given the fast-expanding innovation of self-driving vehicles. Legislative frameworks for AVs, as well as the Automated and Electric Vehicles Act<sup>25</sup> and current efforts by the Law Commission, are steps to filling in the gaps on liability questions.

## **What lessons, if any, can the UK learn from other countries on AI governance?**

Contact details:

*Joshua Krook, Research Fellow, Responsible AI, the University of Southampton. [J.A.Krook@soton.ac.uk](mailto:J.A.Krook@soton.ac.uk)*

*Johanna Walker, Research Associate, Distributed Artificial Intelligence Group, Department of Informatics, King's College London, [johanna.walker@kcl.ac.uk](mailto:johanna.walker@kcl.ac.uk)*

### **Australia**

Australia lacks specific laws and regulations relating to the topic of AI and the usage of autonomous systems, preferring instead to rely on a principles-based approach (described below) to guide usage through norm-setting<sup>26</sup>. This approach has led to the Australian AI Ethics Principles, a "voluntary framework" for businesses, government, and civic society.<sup>27</sup> The AI Ethics Principles are in ongoing development by the Department of Industry, Science and Resources, in consultation with public stakeholders, industry partners and SMEs. The government suggests that organisations may voluntarily choose to follow the AI Ethics Principles to help build trust in their

---

<sup>23</sup> European Commission (EC) (2022). Liability Rules for Artificial Intelligence. [https://ec.europa.eu/info/business-economy-euro/doing-business-eu/contract-rules/digital-contracts/liability-rules-artificial-intelligence\\_en](https://ec.europa.eu/info/business-economy-euro/doing-business-eu/contract-rules/digital-contracts/liability-rules-artificial-intelligence_en)

<sup>24</sup> Whittaker, T (2022). 'Artificial Intelligence Liability: Proposed EU Directive'. <https://blog.burges-salmon.com/post/102hy48/artificial-intelligence-liability-proposed-eu-directive>

<sup>25</sup> The Automated and Electric Vehicles Act (AEV) (2018) UK. <https://www.legislation.gov.uk/ukpga/2018/18/contents/enacted>

<sup>26</sup> Simon Burns, Jen Bradley and Sophie Bogard, 'Facial recognition and artificial intelligence in Australia. Do we need more rules?' Gilbert and Tobin (25 July 2022).

<sup>27</sup> Department of Industry, Science and Resources, Australia's AI Ethics Principles (June, 2021).

products and services, to drive consumer loyalty and to contribute to the positive and beneficial uptake of AI.<sup>28</sup>

The Australian AI Ethics Principles include:

- **Human, societal and environmental wellbeing:** AI systems should benefit individuals, society and the environment.
- **Human-centred values:** AI systems should respect human rights, diversity, and the autonomy of individuals.
- **Fairness:** AI systems should be inclusive and accessible and should not involve or result in unfair discrimination against individuals, communities or groups.
- **Privacy protection and security:** AI systems should respect and uphold privacy rights and data protection and ensure the security of data.
- **Reliability and safety:** AI systems should reliably operate in accordance with their intended purpose.
- **Transparency and explainability:** There should be transparency and responsible disclosure so people can understand when they are being significantly impacted by AI, and can find out when an AI system is engaging with them.
- **Contestability:** When an AI system significantly impacts a person, community, group or environment, there should be a timely process to allow people to challenge the use or outcomes of the AI system.
- **Accountability:** People responsible for the different phases of the AI system lifecycle should be identifiable and accountable for the outcomes of the AI systems, and human oversight of AI systems should be enabled.

As previously stated, these principles are voluntary in nature, and there is currently no foreseeable plan to turn them into laws or regulations that directly impose sanctions on non-compliance.<sup>29</sup> That being said, a wider discussion of the need for AI regulation in Australia has recently started to gain pace due to costly lawsuits, including the Robodebt Scandal (described below), caused by a lack of regulatory information. In 2021, the Australian Human Rights Commission (AHRC) recommended a more hands-on approach to regulation, with laws that regulate the use of artificial intelligence, the establishment of an AI safety commissioner and the regulation of administrative automated decision-making by the government.<sup>30</sup> The government has, in turn, responded by seeking public consultations on the matter. In April 2022, the Digital Technology Taskforce in the Department of the Prime Minister and Cabinet (the Department), opened a call for submissions on an Issue Paper titled Automated Decision-Making and AI Regulation.<sup>31</sup>

---

<sup>28</sup> Department of Industry, Science and Resources, Australia's AI Ethics Principles (June, 2021).

<sup>29</sup> Aimee Chanthadavong, 'Australia's digital minister avoiding legislating AI ethics and will remain voluntary framework' *ZDNet* (27 July 2021); International Bar Association, 'Australia' in *Guidelines and Regulations to Provide Insights on Public Policies to Ensure Artificial Intelligence's Beneficial Use as a Professional Tool* (IBA Alternative and New Law Business Structures Committee, 2021) 28.

<sup>30</sup> Australian Human Rights Commission, Human Rights and Technology – Final Report (2021).

<sup>31</sup> Commonwealth of Australia, Positioning Australia as a leader in digital economy regulation – Automated decision making and AI regulation – Issues Paper (Department of the Prime Minister and Cabinet, March, 2022).

The Department sought advice on how Australia's "regulatory settings and systems can be modernised to ensure they are fit for the digital age and facilitate accelerated and responsible uptake of new technologies."<sup>32</sup> Stakeholders were asked to submit discussion papers on regulatory issues on "bias in automated systems," how to ensure "private information and data remains protected," and how to promote "transparency in ADM [automated decision-making]."<sup>33</sup> Policy recommendations from this paper are likely to occur in 2023, with the potential for subsequent regulatory action in future years.

The government has also established the ARC Centre of Excellence for Automated Decision-Making and Society (AMD+S), a network of universities hosted by Melbourne Law School "to create the knowledge and strategies necessary for responsible, ethical, and inclusive automated decision-making."<sup>34</sup> The Centre has the aim of formulating "world-leading policy and practice in responsible, ethical and inclusive ADM, for governments, industry and the non-profit sectors".<sup>35</sup> This includes the recommendation of policy solutions and legislative proposals.<sup>36</sup> Recommendations by the Centre may help frame future regulations or legislative proposals.

Australia has other relevant laws that inadvertently impact AI. The Privacy Act governs the collection of private information by the federal government and private companies; however,, a recent review commissioned by the Attorney-General suggests that **it is not necessarily fit for purpose when it comes to the challenge of data and AI**. It should be noted that Australia has no bill of rights or formalised system of human rights (such as the right to privacy), outside of state-based laws, such as the Victorian Charter. Explicit rights in the Constitution are limited to the right to vote, the implied freedom of political communication (a narrow version of free speech that does not apply to individuals or groups, only political communication as a whole), and freedom of religion. This hard limit in constitutional law prevents the widespread class action lawsuits against technology companies seen in other jurisdictions that seek to protect individual rights when infringed by artificial intelligence.

The lack of a clear regulatory framework for artificial intelligence has created widespread mistrust in AI systems when used by the government and private companies, resulting in costly lawsuits. In 2021, the federal government settled a 1.76-billion-dollar class action lawsuit with regards to the Robodebt Scandal (officially, Online Compliance Intervention (OCI). The OCI automatically sent debt notices to welfare recipients who were seen to have over-claimed welfare, when compared to average

---

<sup>32</sup> Commonwealth of Australia, Positioning Australia as a leader in digital economy regulation – Automated decision making and AI regulation – Issues Paper (Department of the Prime Minister and Cabinet, March, 2022), p. 1.

<sup>33</sup> Commonwealth of Australia, *Positioning Australia as a leader in digital economy regulation – Automated decision making and AI regulation – Issues Paper* (Department of the Prime Minister and Cabinet, March, 2022), p. 1.

<sup>34</sup> Australian Research Council, 2020 ARC Centre of Excellence for Automated Decision-Making and Society (2020) <<https://www.arc.gov.au/funding-research/discovery-linkage/linkage-program/arc-centres-excellence/2020-arc-centre-excellence-automated-decision-making-and-society>>.

<sup>35</sup> ARC Centre of Excellence for Automated Decision-Making and Society (2020) <<https://www.admscentre.org.au/>>.

<sup>36</sup> ARC Centre of Excellence for Automated Decision-Making and Society (2020) <<https://www.admscentre.org.au/>>.

income calculated from the Australian Taxation Office (ATO) and Centrelink. Previously, discrepancies between average income and claimed income were investigated by humans, at an average of 20,000 investigations a year, of a possible 300,000 discrepancies a year. The OCI automated this process, taking humans out of the loop entirely and sending out debt notices without any supervision, allowing for a far greater number of debt notices per year. In the end, 470,000 debts were seen to have been wrongly issued by the artificial intelligence system. **Taking humans out of the loop can result in costly litigation when automated decisions have real-world consequences** due to errors in systems or datasets.

Uncertainty in the law has also led to government agencies breaching the government's own AI Ethics Principles. In 2020, the Australian Federal Police began using Clearview AI, a facial recognition software that can assist in finding and arresting suspected criminals.<sup>37</sup> Clearview AI was allegedly using illegally scraped data (photos of faces) from social media networks. In 2021, the Information Commissioner and Privacy Commissioner found that Clearview AI had, in fact, "breached Australians' privacy by scraping their biometric information from the web and disclosing it through a facial recognition tool".<sup>38</sup> Data was collected by Clearview AI without consent, and not collected in compliance with the relevant privacy guidelines and regulations. The AFP, by using Clearview AI, itself was in breach of privacy guidelines and the ethical principle of gaining consent for the use of data. Cases like these make it harder for the public to trust government agencies that use AI technologies. Public trust in automated decision-making needs to be entrenched in a rules-based system where the government follows its own rules.

Australian courts have filled in the gaps where AI systems go beyond existing regulations. The federal court, for example, found that AI technology cannot be cited as the sole inventor of a patent, and that a real person has to be credited on a patent application (but AI can be listed as a co-inventor) (*Thaler v Commissioner of Patents* [2021]). Courts have also endorsed the use of AI-assisted review, also known as Technology Assisted Review, for the discovery process of law firms (*McConnell Dowell Constructors (Aust) Pty Ltd v Santam Ltd & Ors* [2016]). For example, a discovery process that would have taken 23,000 hours for humans to complete was allowed to be done with the assistance of technology. This ad-hoc approach to laws on AI will see a slow growth of rules on AI over time. However, the slow rollout of rules prevents business certainty. An established law-based framework on AI could preempt court decisions, allowing businesses, consumers and governments to plan future activities ahead of time with confidence.

## **EU Artificial Intelligence Act (Draft)**

*The EU has defined AI, separately from the underlying technologies*

---

<sup>37</sup> Jake Goldenfein, Australian police are using the Clearview AI facial recognition system with no accountability, *The Conversation* (2020).

<sup>38</sup> Office of the Australian Information Commissioner, Clearview AI breached Australians' privacy (3 November 2021).

The EU recently published its Artificial Intelligence Act. It defines AI as 'software that is developed with one or more of [ a set of techniques and approaches] and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with';<sup>39</sup>

*Risk-appropriate approaches*

A key facet of the Act is the identification of different categories of risk: prohibited, high-risk, limited risk, and low-risk. Rather than a 'one size fits all' approach the EU develops guidance for each category.

*There is a great deal of existing governance support*

The requirement for those involved in operationalising and commercialising AI to conform with the AI Act has led to the creation of multiple governance toolkits, such as the Conformity Assessment Procedure created by the University of Oxford.<sup>40</sup>

---

<sup>39</sup> <https://artificialintelligenceact.eu/the-act/>

<sup>40</sup> Floridi, Luciano and Holweg, Matthias and Taddeo, Mariarosaria and Amaya Silva, Javier and Mökander, Jakob and Wen, Yuni, capAI - A Procedure for Conducting Conformity Assessment of AI Systems in Line with the EU Artificial Intelligence Act (March 23, 2022). Available at SSRN: <https://ssrn.com/abstract=4064091> or <http://dx.doi.org/10.2139/ssrn.4064091>