**Department of Industry, Science and Resources Supporting responsible AI: discussion paper**

**Submission Paper from Linda Shave, FRIM**

**Researcher, Author and Independent Advocate for the Digital Information and Records Management Industry**

**1: Definitions**

**1.1 Do you agree with the definitions in this discussion paper? If not, what definitions do you prefer and why?**

As an Independent advocate for the Digital Information and Records Management industry, I believe that the new definition as outlined in ISO/IEC 22989:20201 **“*AI - capability to acquire, process, create and apply knowledge, held in the form of a model, to conduct one or more given tasks*’’** should be used because it is more apt and more comprehensive for describing or defining Artificial Intelligence.

This definition more accurately represents the ISO/IEC 22989 AI system life cycle model and its mapping with the Organization for Economic Cooperation and Development (OECD) AI life cycle which covers activities such as planning, design, data collection, processing, validation, deployment, operation, monitoring and retirement. This definition also aligns with the functions and activities of capture, use, protection, retention and disposal of data and information assets which are undertaken by information and record management professional practitioners in Government and their agencies.

**2: Potential gaps in approaches**

**2.1 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?**

From a digital, information and records management perspective the privacy by design, data protection by design (DPbD) and safety by design allows digital systems to automatically delete data once it is no longer needed for the specific business purpose. (Page 31).

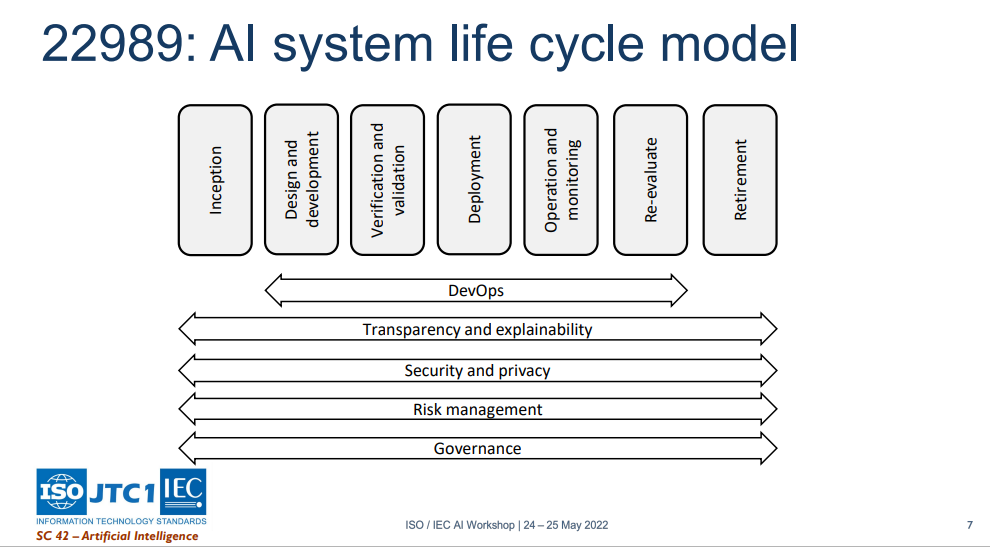
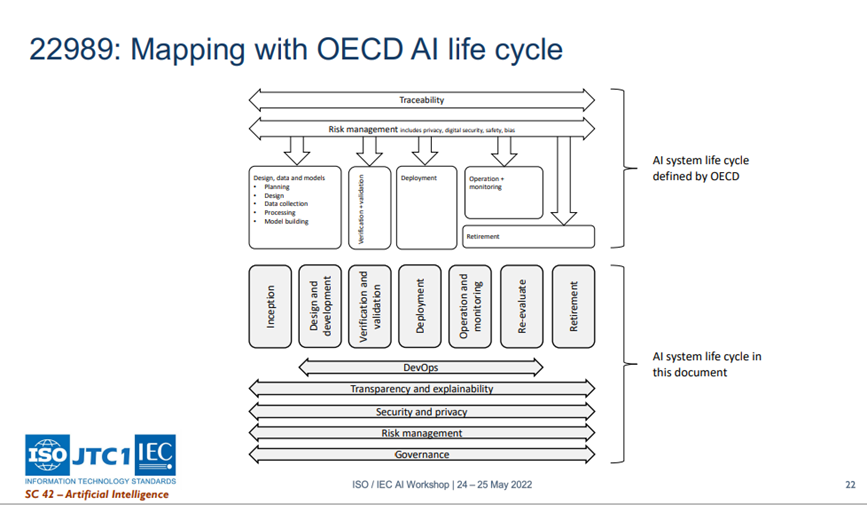
This could be in conflict with the various information and records managements Acts, legislative requirements and best practices at all levels of Government (Local, State and Federal) and their agencies for the capture, use, protection, retention and disposal of data and information assets.

**2.2 Are there any further non-regulatory initiatives the Australian Government could implement to support responsible AI practices in Australia? Please describe these and their benefits or impacts.**

Non-regulatory AI could result in biased development and therefore bias answers, spread misinformation and result in citizen mistrust of the trustworthiness of the App, product or services. Pressure by government to provide non-regulatory initiatives in order to escalate AI innovation and growth could have serious results and prove to be disastrous. It is better to err on the side of caution when dealing with a new and evolving technology. Technologies such as Quantum computing, Large language models (LLMs), Multimodal Foundation Models (MFMs) and relationships with Artificial Intelligence (AI), Machine Learning (ML), Generative AI models (GAI), and Automated Decision Making (ADM) solutions, products and services. It is sensible to regulate or monitor very closely any and all AI innovations.

**2.3 Do you have suggestions on coordination of AI governance across government? Please outline the goals that any coordination mechanisms could achieve and how they could influence the development and uptake of AI in Australia.**

The development of any Government or Agency AI system is in my opinion no different to any other system, product or service project that Government and their agencies undertake. The AI lifecycle should follow the same basic principles of best practices, rules, regulations and all data about the design, development and retirement of the AI application should be documented and follow standard information and records lifecycle management principles. The confidence that protocols are being followed could influence the development and uptake of AI in Australia. Citizens would have confidence that in the result of non-compliance Government and their agencies could respond to any enquiry or litigation as there would be an end to end information and records management life cycle of the project/program. Please also see ISO/IEC 22989 AI system life cycle model below:



**3: Responses suitable for Australia**

**3.1 Are there any governance measures being taken or considered by other countries (including any not discussed in this paper) that are relevant, adaptable and desirable for Australia?**

The UK National AI Strategy outlines in Pillar 3 Governing AI effectively - it states ‘’Ensuring that national governance of AI technologies encourages innovation, investment, protects the public and safeguards our fundamental values, while working with global partners to promote the responsible development of AI internationally’’. [National AI Strategy (publishing.service.gov.uk)](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1020402/National_AI_Strategy_-_PDF_version.pdf).

**4: Target areas**

**4.1 Should different approaches apply to public and private sector use of AI technologies? If so, how should the approaches differ?**

The use of AI in the public/government sector include such areas as traffic flow, healthcare, cyberattack prevention, processing large amounts of data, task automation, smart policymaking and problem solving. The trustworthiness and safety of the Nation is paramount therefore, I believe there is a difference in the use of AI technologies. The private sector develops products and services to make profit. The public sector have a duty of care for all citizens before profit and therefore there is need for different approaches on how to apply AI technology into the public and private sector.

**4.2 How can the Australian Government further support responsible AI practices in its own agencies?**

From a digital information and records management consideration I believe that professional bodies such as RIMPA Global and the Australian Government can support responsible AI practices in agencies. Such partnerships can provide benefit realisation as well as identifying the value of data as an asset not only for the day to day operations but also identifying data and information assets to support AI and Automated Decision Making (ADM).

**4.3 In what circumstances are generic solutions to the risks of AI most valuable? And in what** **circumstances are technology-specific solutions better? Please provide some examples.**

One of the key differences between AI and other traditional technological solutions is that, while the latter are limited to tasks that can be performed by following clear rules defined in advance, AI applications are able to analyse data to identify patterns and make decisions based on them. The same could be said that generic solutions for AI risk management would follow clearly defined rules whilst technology-specific solutions are programmed to follow a defined path. Generic solutions for AI are more valuable when complimented by technology to provide four data-driven outcomes availability, accuracy/reliability, analysis and usability.

**4.4 Given the importance of transparency across the AI lifecycle, please share your thoughts on**

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

Neither ISO/IEC 23894 nor ISO/IEC 42001 provide any guidelines on how to perform an AI system impact assessment. However, better understanding and the adoption of ISO/IEC AWI 12792 Transparency taxonomy of AI systems and ISO/IEC 42005 – AI system impact assessment could aid in mandating transparency requirements and mitigating potential AI risks across the private and public sectors.

**4.5 Do you have suggestions for:**

* **whether any high-risk AI applications or technologies should be banned completely?**
* **criteria or requirements to identify AI applications or technologies that should be banned, and in which contexts?**

I believe that any AI applications or technologies that can be used or utilised for exploiting the vulnerable, disadvantaged, promote woke, mistruths and/or have the potential to harm a person, group or the Nation then they should be banned completely.

The criteria or requirements to identify AI applications or technologies that should be banned and in which contexts should start at the beginning of the design. The AI system should be developed in such a way that the information and logic used to arrive at a decision point can be explained in simplistic terms. Further, it is crucial that developers are conscious of any ‘bias’ creeping into the technical design. The AI’s algorithms, attributes, and correlations must be transparent and open to inspection, it should be clear how decisions are arrived at and also be fully explainable in simple terms. Without these criterions or requirements being met then AI applications or technologies in my opinion should be banned completely.

**4.6 What initiatives or government action can increase public trust in AI deployment to encourage more people to use AI?**

There is a need for greater transparency, responsibility and accountability for protecting datasets, building trust with citizens and safeguarding their safety and privacy. The primary focus of Responsible AI is on assuring the ethical, transparent and responsible use of AI technologies in developing and maintaining AI systems, products and services. Responsible AI can guard against the use of black-box approaches, biased data or algorithms to ensure that automated decisions are justified, explainable and trustworthy. Open ,honest and transparent responsible AI strategies can help build public trust in AI deployment and encourage more people to use AI.

**5: Implications and infrastructure**

**5.1 How would banning high-risk activities (like social scoring or facial recognition technology in certain circumstances) impact Australia’s tech sector and our trade and exports with other countries?**

Remembering that the technology sector is in business for profit and big profits can be achieved from trade and export to other countries. I believe that Government should ban high-risk activities such as social scoring.

AI driven risk scoring such as social scoring is used in multiple areas such as education, finance, insurance, migration etc. Social credit scoring driven by AI and technology is in many cases being enforced, creating a credit score based society. We are now seeing the rise of polygenic risk scoring (PRS) being used to provide a measure of individual disease risk or ethnicity-related genomic scoring thus having the potential to exacerbate ethnically based health discrimination. The introduction in New Zealand of the Equity Adjustment Tool to improve surgical wait list inequities uses an algorithm to rank patients on the basis of ethnicity, geographical location and clinical priority. The New Zealand Equity Adjustment Tool gives priority to Māori and Pacific Island patients, while European New Zealanders and other ethnicities, like Indian and Chinese, are lower-ranked. More recently we have also seen the World Health Organisation Global Digital Health Certificate Network.

A real risk with AI is augmenting and supporting existing human biases such as different ethnicity, cultural attitudes, for example, gender, race and roles in society. An outcome of bias algorithms and bias decision-making coupled with incomplete or inaccurate data has the potential to promote woke, mistruths, fake news and misinformation and runs the risk of accelerating discrimination and segmentation. These examples are in my view high-risk activities.

AI products and services promoted without looking at the overall objectives, impacts and consequences could impact Australia’s tech sector and our trade and exports with other countries if these AI products and services promote discrimination and segmentation and further could harm a person, group or the Nation.

**5.2 What changes (if any) to Australian conformity infrastructure might be required to support assurance processes to mitigate against potential AI risks?**

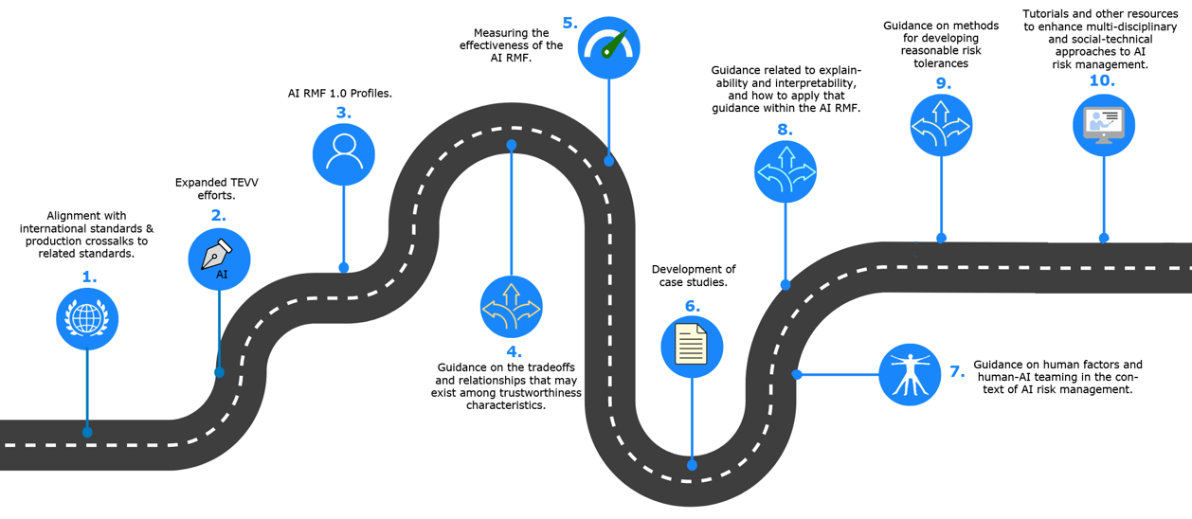
Australia’s standards and conformance infrastructure provides business and consumers with confidence in the goods and service they are developing or using. With the exponential growth for developing AI goods and services there comes greater responsibility and accountability for protecting datasets, building trust with citizens and safeguarding their safety and privacy.

Responsible AI primary focus is on assuring the ethical, transparent and responsible use of AI technologies in developing and maintaining AI systems, products and services. Therefore, the Australian conformity infrastructure needs to be agile to change and develop responsible AI strategies, policies, procedures and risk assessment plans to not only guard against the use of black-box approaches, biased data or algorithms they need to ensure that automated decisions are justified, explainable and trustworthy. Responsible AI strategies need to be in place in order to reduce risk, help build user trust and protect individuals, organisations and national security and privacy.

**6: Risk-based approaches**

**6.1 Do you support a risk-based approach for addressing potential AI risks? If not, is there a better approach?**

Yes, I support a risk-based approach for addressing potential AI risks. The Roadmap for the NIST Artificial Intelligence Risk Management Framework (AI RMF 1.0) goal is to offer a voluntary resource to the organizations designing, developing, deploying, or using AI systems to help manage the many risks of AI and promote trustworthy and responsible development and use of AI systems. See roadmap below. (<https://www.nist.gov/itl/ai-risk-management-framework/roadmap-nist-artificial-intelligence-risk-management-framework-ai>).



**Note: Alignment with international standards and production crosswalks** **to related standards**. (e.g., ISO/IEC 5338, ISO/IEC 38507, ISO/IEC 22989, ISO/IEC 24028, ISO/IEC DIS 42001, and ISO/IEC NP 42005.)

**6.2 What do you see as the main benefits or limitations of a risk-based approach? How can any limitations be overcome?**

The limitations of a risk-based approach is not knowing what risks might be involved nor what evolving technologies might prove to be a risk in the future e.g. Quantum computing. If for example a black box approach has been taken one may not be aware (yet) of what bias or other risks might be lurking under the surface and therefore any risk assessment may not be adequate to assess the risk. The risk will only be identified after the event. Therefore proactive reporting and rectification strategies would need to be enacted.

**6.3 Is a risk-based approach better suited to some sectors, AI applications or organisations than others based on organisation size, AI maturity and resources?**

I believe that all government, agencies and sectors need to utilise a risk-based approach. All AI product, services and applications need to be trustworthy and should incorporate the three core values of being lawful, ethical and robust.

**6.4 What elements should be in a risk-based approach for addressing potential AI risks? Do you support the elements presented in Attachment C? (Attachment C is found on page 40 of the discussion paper)**

In respect to the risk-based elements as outlined in Attachment C found on page 40 of the discussion paper I think that the Australian Government, Department of Industry, Science and Resources should piggyback/adopt work that has already been undertaken to create an AI risk management framework. The following link will take you to the [NIST Trustworthy and Responsible AI Resource Center](https://lnks.gd/l/eyJhbGciOiJIUzI1NiJ9.eyJidWxsZXRpbl9saW5rX2lkIjoxMDEsInVyaSI6ImJwMjpjbGljayIsImJ1bGxldGluX2lkIjoiMjAyMzAzMzAuNzQyOTY5NjEiLCJ1cmwiOiJodHRwczovL2FpcmMubmlzdC5nb3YvSG9tZSJ9.TJPLb6gO2bvecChWZwAdWWJMhxwc7jwkLktblFpSnBc/s/1042468187/br/157132771864-l) (AIRC).

**6.5 How can an AI risk-based approach be incorporated into existing assessment frameworks (like privacy) or risk management processes to streamline and reduce potential duplication?**

I believe there are opportunities for cross mapping of AI risk-based approaches and incorporating them into other existing assessment framework such as privacy, data breach reporting, data protections and information life cycle assessment frameworks. However, one will need to remember that AI data comes from huge amounts of complicated and unorganised, unstructured, structured, semi-structured and sensor data sets in real time. The challenge here is how to validate the quality and identify the value of the data being collected.

The use of intelligent algorithms and machine learning is becoming an essential component of new products and services forming part of the present-day advancements in artificial intelligence. Therefore, AI risk-based approaches need to incorporate risk assessment of the algorithms. A risk-based assessment must not only be agile it should also be mindful that any data collected and stored which has personal identifiers that might identify an individual is removed through the process of anonymisation or de-identification. However, in the era of artificial intelligence this has the potential to become problematic because artificial intelligence systems can be used extremely efficiently to re-identifying people. With algorithms and pattern recognition artificial intelligence systems can be effective at adding data features back into datasets, for example, cross-referencing other available datasets. This in turn could place an organisation at risk of exposing inadvertently information that might identify an individual.

**6.6 How might a risk-based approach apply to general purpose AI systems, such as large language models (LLMs) or multimodal foundation models (MFMs)?**

Firstly, I believe that we need to define what is a ‘general purpose AI system’. Especially, when considering the capabilities of Large Language Models (LLMs) and/or multimodal foundation models (MFMs)such as ChatGPT . ChatGPT uses artificial intelligence and automated decision-making which can cause potential harm as well as having the ability to create deepfakes and algorithmic bias.

**6.7 Should a risk-based approach for responsible AI be a voluntary or self-regulation tool or be mandated through regulation? And should it apply to:**

* **public or private organisations or both?**
* **developers or deployers or both?**

This is a difficult question. However, there is currently much dialog and debate around Artificial Intelligence: innovation, ethics and regulation. Telefónica have suggested a three pillar approach to AI Governance which are global guidelines, self-regulation and a suitable regulatory framework (<https://www.telefonica.com/en/communication-room/blog/artificial-intelligence-innovation-ethics-and-regulation/>). The European Commission is proposing the first-ever legal framework on AI, which addresses the risks of AI. The regulatory proposal aims to provide AI developers, deployers and users with clear requirements and obligations regarding specific uses of AI. At the same time, the proposal seeks to reduce administrative and financial burdens for business, in particular small and medium-sized enterprises (SMEs). <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>.

END