Submission to the Senate Select Committee on Adopting Artificial Intelligence (AI)

May 2024

Submission by:

Dr. Susan Bennett
Principal, Sibenco Legal & Advisory
Founder and Director, InfoGovANZ

Level 26, 1 Bligh Street
Sydney, NSW 2000
E: susan.bennett@sibenco.com
T:+61 28226 8682

1. Introduction

Thank you for the opportunity to provide submissions on the adoption of AI.

While AI brings the potential of enormous societal and organisational opportunities and benefits, it also brings unprecedented risk.

This submission focuses on AI risks and how these can be reduced through clear regulation that is actively enforced and robust organisational governance of AI, data, and information security.

2. Recent trends and opportunities in the development and adoption of AI technologies in Australia and overseas, in particular regarding generative AI

AI in different forms is already in use in many organisations and the AI race is on to seize opportunities and market share by those offering AI products and services across all sectors.

3. Risks and harms arising from the adoption of AI technologies, including bias, discrimination and error

The wide range of risks of harm arising from the adoption of AI technologies have been captured in the U.S. National Institute of Standards and Technology (NIST) draft AI Risk Management Framework: Generative AI Profile (NIST AI 600-1) Initial Public Draft, 29 April 2024 as follows:

1. **CBRN Information:** Lowered barriers to entry or eased access to materially nefarious information related to chemical, biological, radiological, or nuclear (CBRN) weapons, or other dangerous biological materials.
2. **Confabulation:** The production of confidently stated but erroneous or false content (known colloquially as “hallucinations” or “fabrications”).
3. **Dangerous or Violent Recommendations:** Eased production of and access to violent, inciting, radicalizing, or threatening content as well as recommendations to carry out self-harm or conduct criminal or otherwise illegal activities.
4. **Data Privacy**: Leakage and unauthorized disclosure or de-anonymization of biometric, health, location, personally identifiable, or other sensitive data.
5. **Environmental:** Impacts due to high resource utilization in training GAI models, and related outcomes that may result in damage to ecosystems.
6. **Human-AI Configuration:** Arrangement or interaction of humans and AI systems which can result in algorithmic aversion, automation bias or over-reliance, misalignment or mis-specification of goals and/or desired outcomes, deceptive or obfuscating behaviors by AI systems based on programming or anticipated human validation, anthropomorphization, or emotional entanglement between humans and GAI systems; or abuse, misuse, and unsafe repurposing by humans.
7. **Information Integrity:** Lowered barrier to entry to generate and support the exchange and consumption of content which may not be vetted, may not distinguish fact from opinion or acknowledge uncertainties, or could be leveraged for large-scale dis- and mis-information campaigns.
8. **Information Security:** Lowered barriers for offensive cyber capabilities, including ease of security attacks, hacking, malware, phishing, and offensive cyber operations through accelerated automated discovery and exploitation of vulnerabilities; increased available attack surface for targeted cyber attacks, which may compromise the confidentiality and integrity of model weights, code, training data, and outputs.
9. **Intellectual Property:** Eased production of alleged copyrighted, trademarked, or licensed content used without authorization and/or in an infringing manner; eased exposure to trade secrets; or plagiarism or replication with related economic or ethical impacts.
10. **Obscene, Degrading, and/or Abusive Content:** Eased production of and access to obscene, degrading, and/or abusive imagery, including synthetic child sexual abuse material (CSAM), and nonconsensual intimate images (NCII) of adults.

11. **Toxicity, Bias, and Homogenization:** Difficulty controlling public exposure to toxic or hate speech, disparaging or stereotyping content; reduced performance for certain sub-groups or languages other than English due to non-representative inputs; undesired homogeneity in data inputs and outputs resulting in degraded quality of outputs.
12. **Value Chain and Component Integration:** Non-transparent or untraceable integration of upstream third-party components, including data that has been improperly obtained or not cleaned due to increased automation from GAI; improper supplier vetting across the AI lifecycle; or other issues that diminish transparency or accountability for downstream users.'

4. Emerging international approaches to mitigating AI risk

As a Member country of the OECD and a signatory to the Bletchley Declaration, Australia has committed to working co-operatively with other countries to ensure AI design, development, deployment and use is human-centric, trustworthy, responsible and safe. Although AI is subject to existing laws, this submission supports the introduction of AI-specific regulation, which is technology-neutral and risk-based, with a range of sanctions from administrative fines up to and including high-level fines. For regulation to be effective, that is, resulting in human-centric, trustworthy, responsible, and safe AI, it must be supported by active regulatory enforcement and robust organisational governance of AI, data, and information. The emerging approaches to reducing AI risks are considered first from a regulatory perspective, followed by the role of internal frameworks standards and finally by the organisational challenge.

Regulations

On 3 May 2024, the OECD released the updated AI Principles, which includes recommendations for policymakers to 'shap[e] an enabling interoperable governance and policy environment for AI'. This was one of the key updates to the AI Principles at the 2023 Meeting of the Council at Ministerial level, 'underscoring the need for jurisdictions to work together to promote interoperable governance and policy environments for AI, against the increase in AI policy initiatives worldwide.'

The European Parliament adopted the EU's Artificial Intelligence Act (**AI Act**) on 13 March 2024, the first comprehensive AI-specific regulation, which is expected to come into force shortly. The AI Act takes a technology-neutral and risk-based approach to regulating AI. At one end of the spectrum are those **AI systems with minimal risks, which are permitted with no restriction**, at the other end of the spectrum are **specific types of AI with unacceptable risk that are prohibited** – these include systems that manipulate people or exploit people's vulnerabilities, biometric categorisation based on sensitive characteristics and untargeted scraping of facial images from the internet or CCTV footage to create facial recognition databases, and emotion recognition in the workplace and schools, social scoring, predictive policing (when it is based solely on profiling a person or assessing their characteristics). **High-risk systems,** such as, critical infrastructure, education, employment, essential private and public services (e.g. healthcare and banking), migration and border management, justice and democratic processes (e.g. influencing elections) have clear obligations due to their potential harm to health, safety, fundamental rights, environment, democracy and the rule of law. **High-risk systems must assess and reduce risks,** maintain use logs, be transparent and accurate, and ensure human oversight. Citizens will have a right to submit complaints about AI systems and receive explanations about decisions based on high-risk AI systems that affect their rights. **Limited risk systems include general-purpose AI (GPAI) systems** have to meet certain transparency requirements, including compliance with EU copyright law and publishing detailed summaries of the content used for training. GPAI models with high-impact capabilities (e.g. OpenAI's GPT-4) that could pose **systemic risks** have additional requirements, including performing model evaluations, assessing and mitigating systemic risks, and reporting on incidents. Also, artificial or manipulated images (e.g., deepfakes) must be labelled as AI creations.

The Role of Framework and Standards

There has been a gradual evolution from high-level 'ethical AI principles' to formal frameworks and standards development in recent years. Regulations such as the EU AI Act also reference existing and foreshadowed standards. Guidance issued by regulators

and voluntary codes, standards and frameworks are tools organisations can draw on to guide them in implementing appropriate governance policies and processes within their organisation to comply with all applicable laws.

In December 2023, the first international standard on AI was released by the International Standards Organisation **(ISO)** and the International Electrotechnical Commission (IEC) ISO/IEC 42001:2023, *Information technology - Artificial Intelligence - Management System* which is a certifiable framework for an AI Management System to support organisations in the responsible development, delivery, or use of AI systems. On 16 February 2024, Standards Australia announced the adoption of ISO/IEC 42001.

Following President Biden's Executive Order (EO) on the Safe, Secure and Trustworthy Development of AI on 30 October 2023, the National Institute of Standards and Technology (NIST) on 29 April 2024 released draft AI Standards and Risk publications for comment: AI RMF Generative AI Profile (NIST AI 600-1) identifies 12 GenAI risks (listed in section 2 above) and 400 risk mitigation actions; Secure Software Development Practices for Generative AI (NIST SP 800-218A) which is designed to help manage the risks of GenAI; Reducing Risks Posed by Synthetic Content (NIST AI 100-4); Plan for Global Engagement on AI Standards (NIST AI 100-5); and NIST evaluation program to assess GenAI technologies. The first two are guidance documents and are companion documents to NIST's AI Risk Management Framework (AI RMF) and Secure Software Development Framework (SSDF). The publications are initial drafts for public feedback and NIST intends submitting final versions later this year.

On 15 April 2024, the National Security Agency's Artificial Intelligence Security Center (NSA AISC) published the joint Cybersecurity Information Sheet Deploying AI Systems Securely in collaboration with CISA, the Federal Bureau of Investigation (FBI), the Australian Signals Directorate's Australian Cyber Security Centre (ASD ACSC), the Canadian Centre for Cyber Security (CCCS), the New Zealand National Cyber Security Centre (NCSC-NZ), and the United Kingdom's National Cyber Security Centre (NCSC-UK).

While ISO/IEC 42001:2023 and the NIST Framework, Standards and Risk publications are examples of key developments in emerging international approaches to reduce AI risks, **it is important to emphasise their voluntary nature, particularly in the context of a global AI race**. Furthermore, while these standards provide detailed guidance to organisations, they identify the (critical) and very substantial work involved in properly reducing AI risks, particularly the extensive data governance work required for safe AI. The AI governance requirements laid out in standards and frameworks will need to be integrated into existing organisational governance structures within organisations to enable broader risks to be identified and reported to the overarching risk committee and board/governing authority.

The Organisational Governance Challenge

Regulatory compliance is challenging for small, medium, and large organisations – across all sectors. While the desire to innovate using data-driven technology is strong, this submission points to several cases that highlight the struggle large organisations already have to govern and manage regulatory compliance with existing laws, regulatory guidance, risk management frameworks and standards. As demonstrated by the Optus, Medibank, and Latitude data breaches in late 2022 and early 2023, organisations are over-retaining vast volumes of personal information with reports that some people had their personal information (including driver's licences, passports, financial and medical information) disclosed in all three breaches (see for example, article by Emilia Terzon, Latitude customers are furious: some have had data hacked before through Medibank and Optus, ABC News, 18 March 2023). The over-retention of personal information is notwithstanding the current regulatory requirement in Australian Privacy Principle 11.2 that requires entities subject to the *Privacy Act 1998 (Cth)* to keep information for no longer than required. In the government context, the Robodebt scandal and the British Post Office scandal have highlighted the challenges for government in rolling out technology and the serious risks of harm to citizens. Specifically, the case of Robodebt has shone a light on the important role of culture within an organisation and how this can impact the way in which identified problems may not be adequately addressed and

remediated (see Susan Bennett's blog article Five Key Lessons from Robodebt for AI and Technology Projects, March 2024). In the case of the British Post Office scandal, the role of the unhelpful helpline, where struggling users (the Post Masters) of the defective IT Horizon accounting software provided by a third-party provider, Fujitsu, were informed they were the only ones experiencing a problem. Failures of technology integration led to failures in monitoring and reporting under anti-money laundering laws, with a $1.3 billion penalty ordered against Westpac (*Chief Executive Officer of the Australian Transaction Repost and Analysis Centre v Westpac Banking Corporation* [2020] FCA 1538), and a $700 million penalty ordered against CBA (*Chief Executive Officer of the Australian Transaction Reports and Analysis Centre v Commonwealth Bank of Australia Limited* [2018] FCA 930). In both cases, there were technology configuration errors, which were not identified and remediated through risk and compliance, contributing to and resulting in breaches of anti-money laundering laws. The above cases highlight that even in large organisations with dedicated IT, Legal and Compliance functions, integrating data, technology and regulatory compliance is already a significant challenge. Given the significant increase in risks that AI brings, government agencies and organisations will need to ensure they are sufficiently building robust governance processes, particularly around ongoing data governance, information security, regulatory compliance (with an emphasis on privacy compliance) and auditing to minimise these risks.

5. Opportunities to adopt AI in ways that benefit citizens, the environment and/or economic growth, for example in health and climate management

The benefits of AI to citizens need to be considered in light of both positive and negative societal implications, including a significant reduction in labour requirements and the anticipated immense drain on energy availability discussed in 8 below. Given the risks, particularly around implementation (discussed in 4 above), the opportunities to adopt AI need to be strategically considered, appropriately prioritised, and properly resourced with ongoing monitoring and auditing.

6. Opportunities to foster a responsible AI industry in Australia

A responsible AI industry will be assisted by clear AI-specific regulation, supported by education (regulator, education providers, etc.) and active enforcement of that regulation so that the guardrails are clear to AI providers and the organisations using these AI systems. It will be important to enable open AI sandboxes where AI technology providers can demonstrate products to stakeholders, including regulators, and provide opportunities for feedback to minimise risks (where they can be identified) and strengthen privacy and information security. The AI sandbox environment should be developed and be open to all technology providers. Participation by start-ups and small to medium enterprises should, in particular, be encouraged and fostered to support innovation and responsible and safe AI.

7. Potential threats to democracy and trust in institutions from generative AI

Potential and actual threats are already evident. For example, in January 2024, cease-and-desist orders were reported to have been issued against two companies in the U.S. connected with robocalls using AI to mimic President Jo Biden's voice and discourage people from voting in New Hampshire's primary ballot ahead of the Presidential election in November. (See, for example, AP News report Fake Biden robocall being investigated in New Hampshire 23 January 2024).

The European Commission recently announced it has opened formal proceedings to assess whether Meta may have breached the Digital Services Act (DSA). The proceedings will focus on suspected infringements of Meta's policies and practices involving: deceptive advertisements and disinformation; visibility of political content; the non-availability of an effective third-party real-time civic discourse and election-monitoring tool ahead of the upcoming elections to the European Parliament and other elections in various Member States; and the mechanism to flag illegal content (see

European Commission Press Release Commission opens formal proceedings under DSA 30 April 2024).

8. Environmental impacts of AI technologies and opportunities for limiting and mitigating impacts

This submission considers the impacts of AI technologies from first, the organisational ESG responsibility in light of impending mandatory environmental reporting and existing disclosure requirements arising from data storage and data carbon footprint, and second, the foreseeable situation within a decade where the energy demands of data storage are likely to impact the availability of energy for the remainder of the economy and citizens.

With emerging mandatory environmental reporting requirements and the additional focus on ethical use of data and technology, accurate ESG (environmental, social and governance) reporting will become increasingly important. Organisations will need to measure and report on each element of ESG concerning data being collected, generated, used, and stored. These interconnected issues of data collection and processing and regulatory compliance, including the privacy regulatory requirement for data minimisation, that is, keeping personal information for no longer than required (Australian Privacy Principle 11.2), are set out in Dark Data – the risks, costs and ESG (see Susan Bennett blog article, September 2023). This refers to the UK's Digital Decarbonisation Project, which states that 'the data industry is predicted to account for more carbon emissions than the automotive, aviation and energy sectors combined'. The research has designed a toolkit for organisations, including a data carbon ladder to support large data projects (see Thomas Jackson and Ian Richard Hodgkinson (2023) Is there a role for knowledge management in saving the planet from too much data?,Knowledge Management Research & Practice, 21:3, 427-435, DOI: 10.1080/14778238.2023.2192580).

Recent media reports have highlighted the massive energy requirements for data storage and processing in light of the expected uptake in AI use. See for example:

- Europe's hidden energy crisis: Data centers *POLITICO*, 3 October 2022;
- NextDC boss says nuclear should be on table as AI sucks up energy, *The Australian Financial Review,* 11 April 2024 (refers to NextDC $1.3b capital raise for new data centres);
- Data centre builders fight infrastructure projects for heavy cranes *The Australian Financial Review,* 18 April 2024;
- Booming AI demand threatens electricity supply *The Australian Financial Review,* 19 April 2024; and
- Blackstone sees 20pc returns in AI-fuelled data centres, but warns of looming power shortage, *The Australian,* 5 May 2024.

This submission points out that given the significant risks arising from AI, governments and businesses must have a clear AI strategy and adequately understand and provide for the likely ongoing costs, particularly of AI processing and data storage. Furthermore, organisations will need to ensure they have robust AI governance and overall corporate governance to comply with all regulatory requirements, including the requirement to minimise personal information and to report on energy use (ESG) accurately. From this perspective, organisations should be focused on reducing overall data volumes within their control for improved regulatory compliance and to reduce their overall risks and costs, including costs in responding to data breaches, costs of document identification and production in legal proceedings (litigation, regulatory investigations, Royal Commissions), customer access and deletion requests of personal data, and FOI requests.

This submission emphasises that improved data and information governance within government agencies and businesses will enable AI use cases more swiftly. Currently, AI use cases abound within organisations; however, they are impeded by justified data integrity concerns (such as inaccurate data, lack of provenance, and privacy compliance issues). This requires organisations to sufficiently invest in improving data governance so that data can safely be used in AI initiatives and innovation.

Author, Dr. Susan Bennett
PhD(Syd), MBA (AGSM), LLM (Hons) (Syd), LLB(Tas), FGIA, FIP, CIPP/E, CIPT
Principal, Sibenco Legal & Advisory
Founder and Director, InfoGovANZ

Susan is a lawyer with more than 30 years of experience and works at the intersection of data, technology and regulatory compliance, particularly in privacy and governance. For the first 20 years, Susan was a commercial litigator (including as a senior partner at a national law firm) working on large-scale disputes and Royal Commissions. In response to the growing volumes of data stored by organisations, technology was developed in the late 90s and early 2000s to assist in identifying and producing relevant documents required in the 'discovery' process in legal proceedings. For the past decade, this technology has developed to include machine learning capabilities and, more recently, AI. The eDiscovery market, according to Fortune Business Insights, is now valued at $USD15.45 billion and projected to reach $USD40 billion by 2032. Use of eDiscovery technology is expanding to assist in data breach response and FOI and access requests under privacy laws. The prolific growth in data being stored by all organisations, the consequent substantial expense of document production and associated issues (comprising 20-50% of total legal costs incurred in proceedings), and the development of global privacy regulations led Susan to found InfoGovANZ in 2016. InfoGovANZ's mission is to break down information silos across the organisation and bring professionals working across the data and information sphere - Data Privacy, AI and Ethics, Cyber and Information Security, eDiscovery, ESG, Data, FOI, Information Governance, Legal, Records Management, Risk and Compliance - with a multi-disciplinary focus to collaborate and share information governance best practices.

Privacy and Data Protection: the interaction of meta-regulation and information governance
(Doctoral thesis completed 2023)
Abstract: Collection and storage of exponential volumes of personal data give rise to significant opportunities and risks for organisations. The thesis examines the challenge of controlling personal information from the standpoints of the regulator and the regulated organisation. First, the thesis analyses the regulatory design of Australia's Privacy Act 1988 (Cth) and the European Union's General Data Protection Regulation involving the use of principles-based and meta-regulation that devolves the design and implementation of compliance mechanisms to regulated organisations. Second, from the organisational perspective, the thesis examines the challenges for corporate governance when boards must grapple with multifaceted strategic opportunities and risks arising from the intersection of technology, data, and regulation. Based on interview evidence, it develops a theory of effective information governance, which enables data and information to be safely leveraged as a business asset, while ensuring compliance with privacy and other information regulatory and legal requirements. The findings are intended as a practical governance solution to assist organisations in achieving data and privacy meta-regulatory requirements, while pursuing strategic organisational objectives in complex and data-driven operating environments. Available here

AI Risks, Failures and Consequences: Corporate Governance for the AI Era
Dr. Zofia Bednarz and Dr. Susan Bennett (forthcoming publication)
Abstract: Artificial intelligence (AI) tools can bring undeniable benefits for private sector organisations, however they can also lead to significant harms. Focusing on the example of the financial industry, this article explores how these harms translate into risks for organisations in the context of AI applications that have direct implications for consumers. We ask what the use of such AI tools means in terms of risks for companies under current, and emerging, risk and governance frameworks, and what it means for directors discharging their duties under Corporations Act 2001 (Cth) and general law. Drawing on practical examples and recent cases, we examine ways in which the risks of harm can arise and analyse the challenges organisations face when implementing AI governance systems to adequately reduce risks.