Managing AI risks in the Public Sector: A Distributed Digital Leadership Approach.

Boniface Ushaka Adie [a*], *Elizabeth* Valentine [b], *Mary* Tate [c], *Wonhyuk* Cho [d].

[a] Wellington School of Business and Government, Victoria University of Wellington, New Zealand, boni.adie@vuw.ac.nz, ORCID number 0009-0009-9118-9358.

[b] Wellington School of Business and Government, Victoria University of Wellington, New Zealand, lizzie.l.valentine@vuw.ac.nz, ORCID number 0009-0001-4096-6251.

[c] Wellington School of Business and Government, Victoria University of Wellington, New Zealand, mary.tate@vuw.ac.nz, ORCID number 0000-0002-4284-7467.

[d] Wellington School of Business and Government, Victoria University of Wellington, New Zealand, wonhyuk.cho@vuw.ac.nz, ORCID number 0000-0001-7607-6480.

Submitted: 31 January 2025, Revised: 26 March 2025, Accepted: 21 April 2025, Published: 20 May 2025

Abstract. Ensuring organisations are safe from cyber-attacks requires the contribution of every staff member and this also applies to AI risk mitigation. Organisations must assess and implement strategies that ensure AI risk mitigation is not just the responsibility of the cybersecurity team (who are always in short supply), but the entire organisation from the front-line staff who need to classify documents correctly, to the cybersecurity engineers who implement technology solutions to mitigate risks, and top management who drive and implement effective AI strategies, policies and investment prioritisation. In this study, we examine how the distributed digital leadership (DDL) framework can be used to enable agencies across government to mitigate AI risks in the public sector. We conducted semi-structured interviews with cybersecurity practitioners from public sector agencies and determined that AI risk mitigation is indeed everyone's responsibility – spanning people, process, technology and data controls. Using the distributed digital leadership (DDL) framework, we provide actionable suggestions on how collective, and collaborative risk mitigation strategies can be implemented across the public sector by making everyone competent in their respective job role responsibilities. We suggest that a well-aligned DDL can help cushion the skills shortage risks in cybersecurity and the overall management of AI risks in the public service.

Keywords. Cybersecurity, AI risks, Digital Leadership Competencies, Distributed Digital Leadership, All-of-Government.
Research paper, DOI: https://doi.org/10.59490/dgo.2025.957

1. Introduction

The use of Artificial Intelligence (AI) in the public sector has gained momentum over the last couple of years as shown in two recent surveys conducted in the New Zealand public service around the use of AI (DIA, 2025b). The first survey conducted in 2023, was a cross-agency AI survey that focused on the views and uses of AI in the public service. Key takeaways from the survey showed that agencies are increasingly interested in the use of AI, they plan to grow their AI use, knowledge and skills, and they want the benefits that AI provides (efficiency, analysis and service delivery). But they are also concerned about the risks that AI poses including security, privacy and skills shortage (DIA, 2025b). The second survey conducted in 2024 provides an in-depth analysis of how AI is used in the New Zealand public service and the areas of risks that public servants are concerned about. Agencies reported seven areas of risk and barriers for the use of AI in the public service which include: a) skills, capabilities; b) policies, guidance; c) privacy; d) security; e) technical barriers; f) cost, funding; and, g) understanding and support (DIA, 2025b). Similar results were found in other jurisdictions like Canada, the United Kingdom and South Korea where studies of government guidelines for AI use in the public sector have been studied (Beltran et al., 2024). According to Wirtz et al. (2019), the exponential growth of AI has necessitated the need for specialist skills and

competencies to support the implementation of AI in the public sector. Importantly, a common theme that runs through current studies is the lack of AI skills and competencies including in the cybersecurity domain. In addition, the benefits versus risks debate of AI use in the public sector provides paradoxical tensions that are exacerbated by the shortage of cybersecurity talent and competencies across the public sector (Bright et al., 2024; DIA, 2025b). These tensions need to be managed adequately by effective and collaborative AI risk mitigation strategies that will enable the safe use of AI in the public sector.

Managing the tensions requires a new approach that ensures risk mitigation capabilities are embedded in and are contributed from every job/role in the organisation (agency). That way everyone gets to contribute to AI risk mitigation as it is often said 'security is everyone's responsibility'. We argue that by focusing on an All-of-Government (AoG) risk mitigation capability (outcome to be achieved) of the role rather than the number of persons (or skills) that occupy cybersecurity roles, we can achieve better AI risk management outcomes. This means two things: 1) effective risk management as a measurable outcome that is everyone's responsibility, i.e., deliberately and intentionally taking the focus of security risk and mitigation beyond the cybersecurity team, and 2) we must focus on building the skills and competencies in other areas of the business (outside the core cybersecurity team) that contribute to effective AI risk management. For example, data and information management teams should be responsible for developing and enforcing data governance strategies that underpin the use of technology tools in data ecosystems. For example, where technical skills are in short supply, looking more widely across the wider public sector and drawing on the capabilities of different agencies to collectively manage risks that impact the public sector as a whole. Here, if an agency has cybersecurity competencies in AI and has figured out how to safely deploy off-the-shelf AI products, such knowledge can be shared across the sector to benefit agencies without such capabilities. Similarly, effective AI strategies and policies need not be written from scratch by every agency – especially for agencies lacking specialist strategy and policy capabilities. The same can be said of privacy impact assessments, risk assessments, data governance policies, et cetera. These capabilities and artefacts can be shared across the sector, allowing agencies to leverage them within their specific contexts. AI risk management can therefore benefit from an AoG approach underpinned by a distributed digital leadership (DDL) framework, which is the focus of this research. To undertake this, we ask the following research question:

RQ: How can government agencies better manage AI risks across the public sector?

To address this question, we interviewed cybersecurity practitioners in the public sector involved in the day-to-day management of cybersecurity risks in their respective agencies. The rest of this paper is structured as follows. In Section 2 we provide background research covering AI risks and mitigation strategies identified in literature specific to the public sector and in Section 3 we provide a summary of the research methodology. In Section 4 we provide an analysis of our research findings, explaining how DDL can be implemented in Section 5.

2. Background

2.1 AI risks and mitigation strategies in the public sector

Despite the many advantages of using AI, numerous risks abound that may negatively impact the adoption of AI (Beltran et al., 2024) or the realisation of its full potential in the public sector. Knutsen et al. (2024) suggested that implementing AI technologies that are effective and responsibly deployed is a complex undertaking that requires skills and competencies that are in short supply. They also suggested that the benefits and risks of AI are still emerging, making it difficult to manage adequately. Bright et al. (2024) identified some technology risks (e.g. hallucinations, challenges in explainability, and AI bias), economic risks (e.g. dependence on a few corporate organisations for GenAI tools) and the absence of structured skills development programs in AI which may lead to inequities in AI adoption and utilisation. Privacy risks and data leakage are also identified as the unauthorised exposure of personal information or an unintended exposure of sensitive information may negatively impact the adoption of AI tools (Beltran et al., 2024). There is also the issue of technical debt whereby structural redesign challenges and the limited understanding of AI's impact on organisational systems may introduce vulnerabilities or exacerbate existing risks within the technology ecosystem of the organisation (Mergel et al., 2024). Several authors suggest that one of the biggest areas of risk is the lack of skills and competencies in the adoption and management of AI tools in the organisation (Bright et al., 2024; Medaglia et al., 2023; Peretz-Andersson et al., 2021; Persson & Zhang, 2025). This may manifest in the lack of skills for strategizing, leading and governing all types of digital transformation (Valentine, 2016) implementing AI solutions (Mergel et al., 2024), guiding how to appropriately use GenAI tools (Bright et al., 2024), difficulty in integrating AI tools with legacy systems (Persson & Zhang, 2025), over-reliance on private sector expertise (Medaglia et al., 2023) and hesitancy in adopting AI tools (Mergel et al., 2024). The lack of skills and competencies in AI may result in governance, trust and accountability failures for AI systems (Straub et al., 2023) and the ability to adequately identify and mitigate AI risks (Beltran et al., 2024; Knutsen et al., 2024).

In Table 1, we provide a summary of some AI risk mitigation strategies (not an exhaustive list) which shows that skills and competencies in different job roles (not just in cybersecurity) are required to mitigate AI risks.

Tab. 1 – Summary of AI risk mitigations in literature.

Authors	AI risk mitigation strategies
Knutsen et al. (2024)	Enhance cybersecurity programs specific to AI and continuously align AI policy and strategy, learning, collaboration and information sharing.
Bright et al. (2024)	Provide National guidance, training and education, and organisational support, for the use of GenAI. Use open-source solutions like BLOOM to democratise access to GenAI and reduce dependency on large corporations.
Mergel et al. (2024)	Monitor AI privacy concerns, enhance user engagement, and develop an AI maturity framework. Promote peer learning and establish intergovernmental forums to share knowledge about AI implementation.
Beltran et al. (2024)	Ensure adherence to data protection mechanisms and compliance with privacy regulations for AI. Implement cybersecurity standards and risk assessments for AI. Develop explainable AI models and transparency initiatives. Undertake workforce training to develop AI literacy and encourage critical evaluation of AI-generated content.
Medaglia et al. (2023)	Codify ethical AI standards and regulations and monitor their enforcement, fostering data availability and accessibility through centralized data platforms.
Persson and Zhang (2025)	Modernise IT infrastructure to support AI technologies, establish data quality, model independence, adopt ethical frameworks to address societal impacts and prevent bias in AI systems.
Peretz-Andersson et al. (2021)	Foster interdisciplinary research to provide clear definitions and frameworks for AI transformation. Implement top-down and bottom-up strategies for AI transformation, and establish governance frameworks that address responsibility, ethics, and societal communication.

Consistent across these risk mitigation strategies (and as shown in the New Zealand survey), is the need for competencies in AI risk management and mitigation strategies which should include people, process, technology and data controls (Duong et al., 2024) that would enable the successful development and adoption of ethical AI in the public sector (Straub et al., 2023).

People-oriented controls are centred around end users (Duong et al., 2024) and the skills they need (Ferdynandus et al., 2024) to use AI tools effectively. This requires training and education for all management and staff in the organisation centred around security awareness, privacy awareness, AI strategy awareness, et cetera. Risk awareness and mitigation at this level enable the proper classification of data, setting document permissions, management of personal information, et cetera. Process controls ensure that proper business processes and procedures (Duong et al., 2024) have been put in place to mitigate risks and are enforced. For example, ensuring that privacy impact and data assessments are conducted at the start of an AI project implementation rather than at the end and that AI guardrails are identified and implemented from the start. Technology controls relate to the underlying AI technology infrastructure (Ferdynandus et al., 2024), ensuring that technical settings are configured correctly, security and privacy are present by design, AI integrations to legacy systems are managed properly and that security controls are monitored, enforced and operationally effective. Finally, data controls have to do with technical and non-technical controls that govern the use of organisational data. These controls could range from having a holistic data strategy, data governance, data quality enforcement, data loss prevention, data lifecycle management (Duong et al., 2024), et cetera. AI risk mitigation draws on the collective responsibilities and capabilities of data management, privacy, cybersecurity, technology and strategic job roles (Sattlegger & Bharosa, 2024) within the organisation and across the public sector (Knutsen et al., 2024) for effective risk management. In other words, an All-of-government (AoG) approach to AI risk management is required and introduced in the next section.

2.2 AoG approach to cybersecurity risk management

Mature e-government requires both vertical and horizontal integration of information systems across multiple government agencies and departments that provide similar aspects of a shared service to citizens (Layne & Lee, 2001). For example, there can be horizontal integration between government departments that provide social services to citizens which may entail the provision of financial assistance administered by the tax department. This joined-up (all-of-government, or whole-of-government) approach to digital services provides a one-stop-shop for citizens accessing government services through digital channels and enables improved efficiency, enhanced

service delivery (or new service delivery channels e.g., AI) and value creation for citizens. Thus, digital transformation at this stage begins to deliver digital government benefits (Lindgren & Van Veenstra, 2018) because public sector agencies are working together towards a shared goal (Janowski, 2015) – e.g. AI risk mitigation, efficiency gains and cost reduction.

To fully maximise the potential of new digital technologies, public sector agencies need to rethink their institutionalised assets, structures, leadership orientations (job roles accountabilities), skills and competencies because such transformations trigger significant changes that must be properly managed to achieve intended outcomes (Tangi et al., 2021). The alignment of job roles and accountabilities to competencies underpins the success of digital transformation and digital government in the public sector and provides the basis by which leadership competencies can be matched to strategy (Leblanc & Gillies, 2005). In New Zealand for example, the Strategy for a Digital Public Service (NZ SDPS) sets the strategic direction for the digital transformation of the public sector with digital leadership as a key aspect of the strategy. In the strategy, it is stated that digital leadership among other things aims to identify and grow talent at all levels that are diverse and multi-disciplinary, to deliver public sector-wide results – including ensuring cybersecurity resilience and support for government agencies (Adie et al., 2024b; DIA, 2020; Ushaka Adie et al.). Further to the NZ SDPS, is a service modernisation roadmap which is a 3-year programme of customer-focused digital initiatives from across the public sector requiring agencies to implement initiatives that will benefit the public sector as a whole (DIA, 2024b). Again, this reflects the need for more distributed digital leadership in the public sector (Adie et al., 2024a).

2.3 Distributed digital leadership for cybersecurity risk management

We define distributed digital leadership (DDL) as ***the intentional alignment of job-role accountability of digital leaders and the adoption of their shared capabilities for digital government outcomes***. We adopt this working definition as it best describes the DDL framework conceptualised by Adie et al. (2024a) and suggests that this framework can be adapted to support the AoG management of AI risks. This is especially important given the shortage of cybersecurity competencies (DIA, 2025b) in the public sector and the requirement for other job roles to contribute to AI risk mitigation strategies. According to Khisro (2025, p. 1859), "*…AI capability encompasses not only expertise but also various technical and non-technical components [competencies] that are necessary for effectively developing and utilising AI technologies*". The DDL framework demonstrates how both technical and non-technical roles must align (both within an agency and across the public sector) for the effective deployment and mitigation of AI risks in the public sector.

3. Research Methods

We adopted a qualitative research methodology to understand the views of cybersecurity practitioners on AI risks and mitigation strategies in the public sector as they are involved in the day-to-day management of cybersecurity risks (Myers, 1997).

2.1 Data collection

We conducted a semi-structured interview with 21 cybersecurity professionals across four different agencies including the lead agency for digital transformation in the New Zealand public sector. Semi-structured interviews are used when the study would benefit from an open framework and where they can provide more useful insights from a small sample size of interviewees (Pathak & Intratat, 2012). Using semi-structured interviews allowed us to talk about the main themes of AI risks and mitigation strategies, while also allowing the direction of the interview to be shaped by the participant's practical understanding of these (O'Keeffe et al., 2016). We used purposive sampling (Miles Matthew et al., 2020) to recruit our interview participants because we: a) wanted to talk to cybersecurity practitioners involved in the day-to-day management of AI risks in the public sector; and, b) wanted to talk to agencies who are engaging in DDL practice as suggested by Adie et al. (2024a). Those interviewed included three Chief Information Security Officers (CISO), five cybersecurity managers, and thirteen team members comprising cybersecurity architects, analysts and advisors across the four government agencies. We asked these cybersecurity practitioners about their views on the use of AI in government – focusing on AI risk mitigation strategies.

2.2 Data analysis

Interviews were recorded, transcribed and coded into themes using a thematic analysis approach (Braun & Clarke, 2012) similar to (Kumar et al., 2025) beginning with open codes, axial codes (themes) and selective codes (dimensions) relevant to AI risk mitigation strategies. We also provided suggestions on risk mitigation categorisation showing that mitigating strategies span people, process, technology and data controls (Duong et al., 2024). Although not included in our final codes, we highlighted in our discussion section, some of the areas of concern that cybersecurity practitioners have regarding the use of AI in government. Finally, using the DDL framework proposed by Adie et al. (2024a), we provide a practical analysis of how AI risk mitigation strategies can be implemented using the DDL approach. Note that for privacy reasons and given the relatively small cybersecurity

community, the identities of interview participants were anonymised using pseudonyms (see Appendix A).

The DDL framework is underpinned by the following principles. First, everyone in the organisation contributes to the digital transformation outcomes of the organisation (e.g. cybersecurity is everyone's business) with varying job role accountabilities and should be considered digital leaders in their own right. Second, every job role in the organisation requires digital leadership competencies at different levels of expertise depending on job role accountability. Third, for organisational outcomes to be met (e.g. safe use of AI), there must be an intentional alignment of job role accountability of digital leaders and adoption of shared competencies which are the building blocks of capability (McClelland & Boyatzis, 1980; Peppard & Ward, 2004). Finally, given the lack of competencies in niche areas, agencies across the public sector can contribute to and share the collective capabilities across the sector based on each agency's areas of expertise and digital leadership.

4. Findings and Discussions

We categorised AI risk in government into five broad risk areas namely, data risks, privacy risks, data sovereignty and jurisdictional risks, technology risks and AI strategy and adoption risks as shown in Appendix A. Data risks are predominantly around data leakage and exposure of government data through the use of AI tools. A senior cybersecurity architect puts it this way:

> "*There's lots of instances where people have taken their work and put it into a large language model, AI, and said do my job for me, not realising that it's the internet. That thing[data] now is out there for everyone to see. you've just data breached yourself.*" (PCA6).

Privacy risks are associated with privacy breaches and the exposure of personal information to unauthorised parties. Cybersecurity practitioners are concerned that AI tools would make it easier for privacy breaches to occur. Technology risks are associated with the underlying technical infrastructure and mechanisms of how AI works giving rise to bias, hallucinations, or vulnerabilities in AI models. Finally, AI strategy and adoption risks have to do with inadequate leadership and governance oversight of the adoption and implementation of AI tools resulting from a lack of AI strategy, regulation or guardrails. It should be noted that cybersecurity practitioners are also concerned about environmental and sustainability issues with the adoption of AI. Some of these concerns include the cost of adoption of AI for agencies (PCA2) and the environmental impact of vendors' hyperscale data centres that are required to run AI (see LCA2, PCA2, PCA3 in Appendix A). Whilst these concerns are not necessarily cybersecurity risks, if not managed, may impact the adoption or perception of AI in government as Toll et al. (2019) suggested. These risks are mitigated by strategies categorised into four broad areas including AI strategy and policies, AI best practice standards & guidelines, AI risk assessments & security controls and finally, AI awareness & education as shown in Figure 1 below.

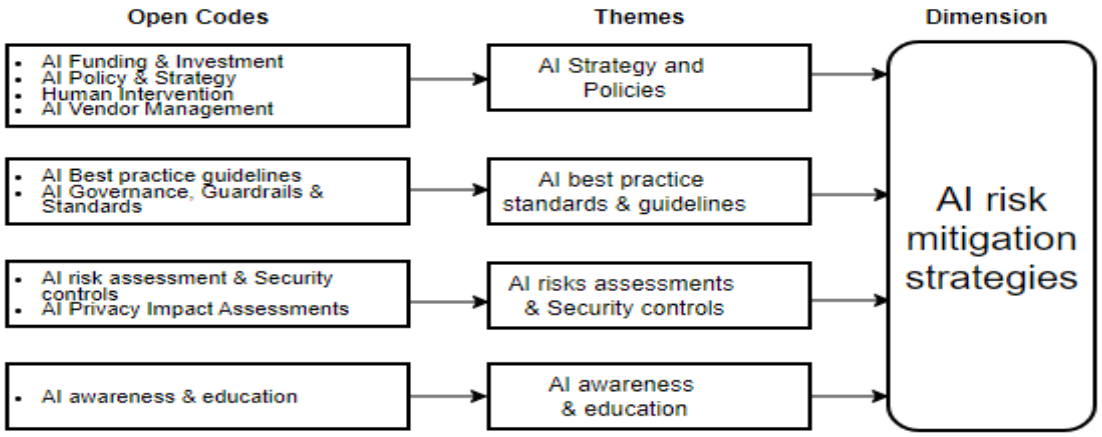

Fig. 1 – AI risk mitigation strategies.

AI strategies and policies primarily relate to people and process (and some governance) controls. They set the stage for how AI is to be used in the organisation and should include requirements for funding and investment, appropriate use, risk management, vendor management and guardrails that must be adhered to. AI strategy and policies need to be socialised, monitored against standards and measures and supported by top management for implementation and adherence to be effective.

AI best practice standards & guidelines touch on process, technology and data controls and are technical documents that are targeted at those tasked with the implementation of AI tools within the organisation. These standards and guidelines could come from technical subject matter experts within the organisation, vendor support documentation or public sector-wide compliance regulations and legislation (Beltran et al., 2024). A cybersecurity practitioner puts it this way "those foundational guardrails really need to be documented or taught about before we go off. Because once you've got those, they can apply to any of those [AI] solutions. [We should be]

prioritizing those guardrails over implementing AI use cases" LCA1. Again, competent monitoring and support are essential.

Conducting AI risk assessments & enforcing security controls is primarily a process control but touches on other areas as risk assessments also involve people, technology and data risks. Risk assessment is probably the most significant AI risk mitigation strategy as signified by the following statements from cybersecurity practitioners – "[conduct] risk assessment, and make sure we have related controls in place which are specific to AI" (LCA5), "the risk appetite of different organisations is quite different" (PCA1), "Do we know what data we are giving away or sharing? Do we know who's going to have access to it?" (PCA7), "Someone needs to monitor and track the AI [tool]. It needs all of those deep level security assessments. regular assessments and regular assurance" (PDT1). Risk assessment for AI should not only be limited to the underlying technology, but it should also cover privacy impact assessments (LDT3, PDT1) and data risk assessment (PCA4) and it should assess the organisational capability and readiness for AI adoption based on their risk appetite and the operational effectiveness of their people, process, technology and data controls.

AI awareness & education are people and process controls that should be tailored to all job roles within the organisation. From general AI awareness programs for all staff to technical training programs provided to those in specialist roles like data management, cybersecurity and architecture. AI training and awareness programs must be determined at the start, needs-based, continuous and updated to keep pace with the ever-changing landscape of the technology (PCA2, PDT2). Some AI vendors offer training and hands-on support in addition to technical documentation and guidelines that can be made available as part of their contractual obligations and vendor support. Microsoft, for example, provide Copilot training programs and technical documentation for the use of Microsoft Copilot in an organisation (Microsoft, 2025d).

These risk mitigation strategies require the deployment of technical and non-technical capabilities (Khisro, 2025) both within an agency and across agency boundaries for public sector-wide AI risk mitigation. However, for DDL to work optimally, the capability of agency leaders to lead, monitor and direct transformation involving AI becomes pivotal (Adie et al., 2024a; Valentine, 2016) as discussed in the next section.

5. Distributed digital leadership approach to AI risk mitigation

Managing cybersecurity risks in the public sector needs a collaborative approach and AI risk mitigation is no different. If cybersecurity is everyone's responsibility – from frontline staff to senior executives - everyone is required to at the very least know how to identify phishing emails and ensure personally identifiable information (PII) is handled according to the agency's privacy policies. Working collaboratively for risk mitigation within the organisation and across the public sector is at the heart of DDL. Cybersecurity practitioners said the following "*It doesn't matter what role you're in, but everyone has been employed because they've got something to bring to the table, so …it is important to be able to work collaboratively to get the best outcome*" (PDT5), "*There is no way that everybody should be an expert and a master in everything*" (LCA2), "*I don't think one person can do it all and I think exactly like you said, a team should play to its strengths and so an organization should play to its strengths*" (PCA5). Whilst specialist cybersecurity, privacy and data management teams ensure that AI risk mitigation strategies are in place and operating effectively, everyone in the organisation contributes to risk mitigation in one way or another. This is shown in Figure 2 which examines common AI risks by type and provides a 2PTD (people, processes, technology and data) approach to mitigation across each risk.

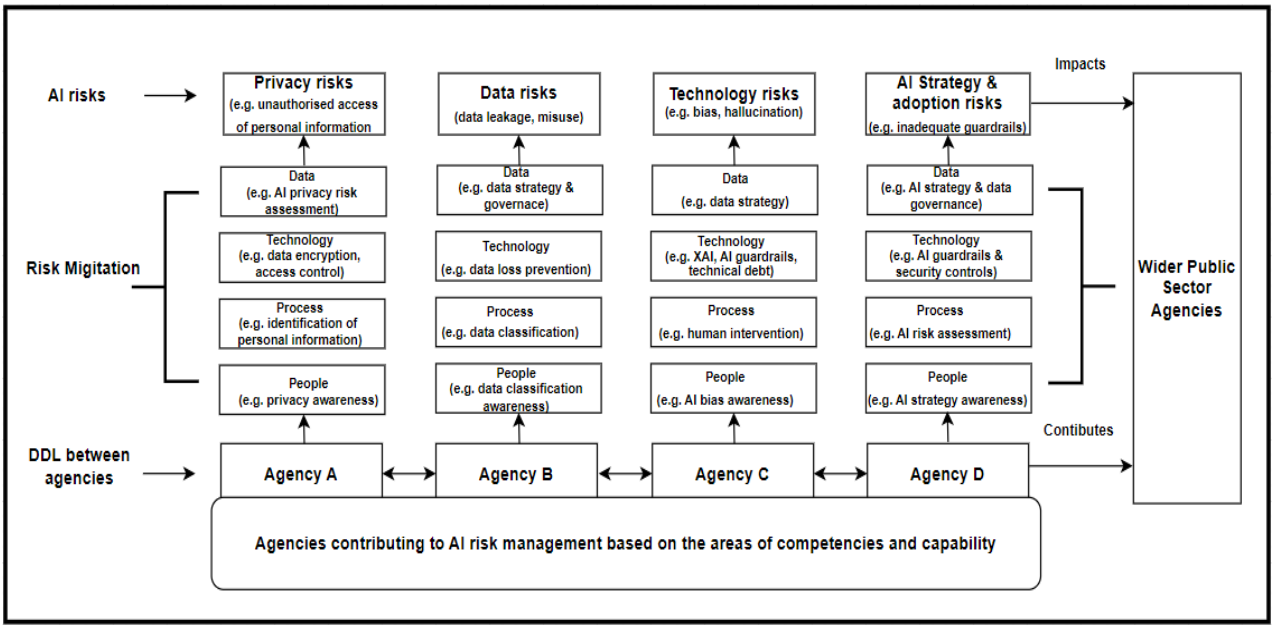

Fig. 2 – Distributed digital leadership approach to AI risk management.

AI risks can be categorised into privacy risks, data risks, technology risks and AI adoption risks. Privacy risks relate to unauthorised access or disclosure of personal information, and these can eventuate when AI agents and models are able to access and surface personal information without prior authorisation and consent. Data risks include data leakage, misuse and exposure of sensitive information especially when effective data governance and access controls are not in place. Technology risks are risks inherent in the use of AI. These risks include AI bias, hallucinations, large language model (LLM) poisoning, et cetera. Finally, AI strategy and adoption risks are risks arising from poor adoption and use of AI in an organisation such as deploying AI without proper understanding of organisational context, use cases and risk mitigation strategies.

As introduced in Section 2, AI risk mitigation strategies must include the 2PTD (people, processes, technology and data) security controls (Duong et al., 2024) with clear roles and responsibilities for job roles across the organisation (Sattlegger & Bharosa, 2024) and the wider public sector – hence a DDL approach. As shown in Figure 2, taking a DDL approach has the potential to positively impact and contribute to improved outcomes across the public sector. There is a many-to-many relationship between risk mitigation controls in the 2PTD framework. For example, AI best practice standards & guidelines are primarily technology controls but are also related to people and processes because technology controls are enforced by technical teams (people) and do require documented processes (process) for continuous improvement and assurance. Similarly, conducting AI risk assessment is a process control but those assessments would usually include people, technology and data risks.

These risk-mitigating controls require capabilities from various teams in the organisation and across the wider public sector to support agencies within the sector, especially smaller agencies that may not have AI risk management capabilities in-house. In Table 2, we provide a summary of how a small sample of different roles contributes to AI risk mitigation using Microsoft Copilot. Microsoft 365 Copilot coordinates large language models (LLMs) which include pre-trained models like Generative Pre-Trained Transformers (GPT-4), integrated into Microsoft 365 productivity apps (e.g. Word, Excel, Teams, et cetera), providing real-time intelligence that enables users to complete tasks more efficiently and enhance their productivity (Microsoft, 2025c). Agencies in the New Zealand public sector that use Microsoft products would most likely have Microsoft 365 Copilot included in the enterprise licenses.

Tab. 2 – Job role contribution to AI risk mitigation using a DDL approach.

Job/role	Contribution to AI risk mitigation	AI risk mitigation strategy
Senior leaders and management teams	Senior management should develop and enforce the implementation of AI strategies, goals, measures of success and governance mechanisms (Valentine, 2016) that govern the use of AI in the organisation including the use of GenAI tools like Microsoft Copilot.	AI strategy and policies. 2PTD: primarily relating to ***people and process*** controls. Example: AI strategy and policies.
Technology, Architecture & Operations	Support the development and enforcement of AI strategy. Ensure alignment with the wider technology and organisational strategies. For Microsoft Copilot,	AI best practice standards & guidelines.

teams	Technology, Architecture & Operations teams should ensure implementation requirements align with organisational technology and AI strategies. Copilot prerequisites, permissions, integrations and guardrails are set up properly (Microsoft, 2025d) and are continuously managed.	2PTD: primarily relating to ***technology*** controls but touches on ***people*** and ***process.*** Example: AI strategy enforcement, AI guardrails, AI configuration, management of technical debt and technology integrations.
Data & Information management teams	Data teams design and govern organisational data strategy that stipulates data governance and management practices including the use of AI. Information management practices should include proper permissions management for sensitive documents, data management lifecycle, enforcement of data owner responsibilities et cetera. For Microsoft Copilot, this would mean ensuring that connections to organisational data sources are done following data governance settings, data risk assessment or regulatory obligations. For example, connections could be prohibited to certain data sources with sensitive or personal information.	AI best practice standards & guidelines. 2PTD: primarily relating to ***data*** controls but touches on ***people process,*** and ***technology.*** Example: Data strategy, data governance, data quality, data loss prevention policies, and data classification.
Cybersecurity teams	Cybersecurity practitioners undertake risk assessments on AI tools and underlying technology risks. They also manage and assess the operational effectiveness of tools that enforce security controls, e.g. access control, data loss prevention, et cetera. For Microsoft Copilot, Cybersecurity would ensure that adequate guardrails (Microsoft, 2025a) are in place (across people, processes, technology and data controls) following organisational risk appetite.	AI risk assessments & security controls. 2PTD: primarily relating to ***process*** controls but touches on ***people, technology and data.*** Example: AI risk assessments and security controls audit.
Privacy teams	Privacy teams undertake privacy impact assessments (PIA) on AI tools and ensure that appropriate controls are in place to mitigate against unauthorised access to personal information. This would mean conducting a PIA for Microsoft Copilot to ensure that access to personal information is done following privacy regulations and organisational requirements (Microsoft, 2025b).	AI (privacy) risk assessments & security controls. 2PTD: primarily relating to ***process*** controls but touches on ***people, technology and data.*** Example: AI privacy impact assessment and enforcement of privacy regulations.
Business analysts	Business analysts document business requirements for AI (Microsoft Copilot), ensuring it aligns with technology and AI strategies. Business Analysts could also help design, document and implement people and process controls including capability and competency analysis as well as establishing strategy-aligned measures of value, uptake and usability.	AI strategy & policies, AI awareness & education. 2PTD: primarily relating to ***people and process*** controls. Example: AI requirements gathering, process controls enforcement, and documentation.
All staff	Understand privacy risks, and data classification for Microsoft Copilot, AI awareness, data strategy, organisational strategy, et cetera. Individual job-role accountabilities are specified and performance standards established,	AI awareness & education. 2PTD: primarily relating to ***people and process*** controls. Example: Privacy awareness, proper data classification, AI bias & strategy awareness.

Every staff member in the organisation should be trained on AI- skills relevant to their job role contribution, with training including AI risks and mitigation strategies to ensure the responsible use of AI tools (Wirtz et al., 2019). As shown in Table 2, AI risk mitigation requires contributions from both technical and non-technical teams within

the organisation with strong oversight from senior leaders and their management teams. Senior leaders' understanding of data and digital transformation is a key digital leadership competence (Ushaka Adie et al.) that should be deployed in the creation of AI strategies that support organisational objectives and minimise risks, and this responsibility should not be left to the IT shop or cybersecurity teams alone.

AI risk mitigation capabilities can also be drawn from the wider public sector. It shouldn't matter what agency AI specialist skills and competencies are located in. Every other agency within the public sector should benefit from the capability they provide, thus, demonstrating DDL at an AoG scale. For example, if Agency A has Privacy Officers with specialist skills in privacy impact assessment for AI tools (e.g. Microsoft Copilot), their knowledge should be shared with the wider sector including artefacts that will enable other agencies on their privacy assessments. Technical teams in agencies that have safely deployed AI tools could share design artefacts or configuration details with other agencies deploying similar AI tools. Security risk assessments could also be shared to enable smaller agencies with no in-house capability to decide on their AI implementation strategy.

Agencies within the public sector can leverage specialist capabilities (across AI strategy, privacy, data, technology, et cetera), create synergies, and support each other to effectively manage AI risks in the public sector. In New Zealand, for example, this is demonstrated in the service modernisation roadmap by the digital leadership of the AoG System Leads for data, digital, procurement, cybersecurity, et cetera (Adie et al., 2024a; DIA, 2024a, 2024b). In the service modernisation roadmap, we find the following examples of how DDL is demonstrated. A) the Government Chief Data Steward (GCDS) leads the initiatives "using data standards more" and the "plan for administrative data pipeline" which both require working with key agencies to identify opportunities for implementing data standards across the sector and developing a plan to build a high-quality administrative data pipeline to provide improved data for service delivery. B) the Government Chief Information Security Officer (GCISO) leads the "Cybersecurity standards", "Cybersecurity guidance" (NCSC, 2025), "Threat and vulnerability management support", and "Secure infrastructure" across government. C) the Government Chief Digital Officer (GCDO) leads the "Government AI search assistant" initiative which involves developing an AI virtual assistant to search across government websites to make it easier for New Zealanders to find the information and services they need. The GCDO also leads the "AoG design system and service design" and "AoG AI work programme" contributing to and supporting agencies to use AI for its benefits while managing the risks (DIA, 2025a). They also lead the "Digital workforce planning" initiative which ensures that the public service can meet current and future service delivery needs. All of these agencies contribute to cybersecurity risk management and AI risk mitigation in the public sector.

We note that our findings and contributions can be applied to jurisdictions with similar public sector structure and democratic settings as New Zealand's, however, where a System Leads approach does not exist, inter-agency knowledge sharing and collaboration should still be encouraged for faster adoption of AI and risk management across the public sector given they all work toward the same outcomes – efficient, trust-worthy service delivery to citizens.

6. Conclusion

There has been a lot of focus on AI risks in the public sector but less on how these risks can be mitigated by leveraging capabilities from different teams within an agency and across the public sector. Our study focused on AI risk mitigation strategies and suggested that Distributed Digital Leadership competencies (Adie et al., 2024a) can be implemented as a risk mitigation strategy especially given and lack of skills and competencies as a risk factor for AI adoption. New technologies like AI provide benefits to the public sector (Bright et al., 2024; Toll et al., 2019) and also risks and barriers (Millan-Vargas et al., 2024) that need to be managed and governed properly (Ganapati & Desouza, 2024). Global Cybersecurity skills shortage (Franco et al., 2024) necessitates a DDL approach to AI risk management in the public sector allowing agencies to share the capabilities (Franco et al., 2024) required for effective risk mitigation of new and emerging technologies.

We have demonstrated that by focusing on the AI strategy, policy and risk management capabilities, agencies and teams (across data, strategy, privacy, business analysis, technology & architecture, et cetera) can contribute to AI risk mitigation based on their job roles accountabilities and capabilities, explaining how DDL is applied within an organisational setting. Also, by leveraging AI risk mitigation capabilities of different job roles both within an agency, and across the sector, government agencies can better manage AI risks across the public sector – thus, addressing our research question.

This study provides two contributions. First, from a theoretical perspective, AI risk management using a DDL approach illustrates the efficacy of AoG approaches to service delivery and technology management across the public sector. Our study emphasises the need for cross-sector collaboration not just for service delivery but for efficiency gains, effective resource utilization (during skills shortages) and for cost savings. Second, from a practical perspective, our study provides practical examples of how distributed digital leadership capabilities across the public sector can be deployed for effective AI risk management. High cost, sustainable implementation of AI,

financial resources and technological capabilities were identified as some of the barriers to the adoption of AI in the public sector (Millan-Vargas et al., 2024). Our study provides practical examples of how capabilities from different job roles across the public sector could be leveraged to address these barriers. A Cybersecurity practitioner captures it this way "*and I do think from a leadership side if we can share competencies across government departments there must be huge savings to be made*" (PDT6). Managing AI risks is indeed everyone's business (responsibility) and distributed digital leadership is one way to operationalise it.

6.1 Limitations and Future Research

Our research showed that DDL can be used to address AI risks and mitigations in the public sector especially as it relates to enabling the distributed deployment of capabilities both within the agency and across the public sector. However, the implementation of DDL may be faced with some challenges such as the lack of annual capability analysis across the public sector, information sharing between agencies and other organisational challenges that need to be addressed which have not been covered in this study. Future research is required to address these challenges and to identify cross-agency factors, processes, and organisational settings that will support the effective implementation of the DDL approach across agencies.

Acknowledgement

- **Contributor Statement**: Author 1: conceptualisation, methodology, formal analysis, writing – original draft. Author 2: supervision, writing – review & editing. Author 3: supervision, writing – review & editing. Author 4: supervision.
- **Use of AI**: During the preparation of this work, the authors used notta.ai for interview data transcription. After using this tool, the authors reviewed, edited, made the content their own and validated the outcome as needed, and take(s) full responsibility for the content of the publication.
- **Conflict Of Interest (COI)**: There is no conflict of interest.

References

Adie, B. U., Tate, M., Valentine, E., & Cho, W. (2024a). Conceptualising Distributed Digital Leadership in the Public Sector. ACIS 2024 Proceedings. 44,

Adie, B. U., Tate, M., Valentine, E., & Cho, W. (2024b). Digital Leadership Competencies for Digital Government: Insights and Implications from New Zealand Government Agencies. Proceedings of the 25th Annual International Conference on Digital Government Research,

Beltran, M. A., Ruiz Mondragon, M. I., & Han, S. H. (2024). Comparative analysis of generative AI risks in the public sector. Proceedings of the 25th Annual International Conference on Digital Government Research,

Braun, V., & Clarke, V. (2012). Thematic analysis.

Bright, J., Enock, F., Esnaashari, S., Francis, J., Hashem, Y., & Morgan, D. (2024). Generative AI is already widespread in the public sector: Evidence from a survey of UK public sector professionals. Digital Government: Research and Practice.

DIA. (2020). Strategy for a Digital Public Service. D. o. I. Affairs. https://www.digital.govt.nz/digital-government/strategy/strategy-summary/strategy-for-a-digital-public-service/

DIA. (2024a). Digital Government Leadership. https://www.digital.govt.nz/digital-government/leadership/

DIA. (2024b). Service Modernisation Roadmap. DIA. https://www.digital.govt.nz/digital-government/strategy/strategy-summary/service-modernisation-roadmap

DIA. (2025a). Public Service AI Framework. DIA. https://www.digital.govt.nz/standards-and-guidance/technology-and-architecture/artificial-intelligence/public-service-artificial-intelligence-framework

DIA. (2025b). Research - AI in the public service. DIA. https://www.digital.govt.nz/standards-and-guidance/technology-and-architecture/artificial-intelligence/research-ai-public-service

Duong, D., Sarbazhosseini, H., Alsheikh, M. A., & Ma, W. (2024). Conceptualising People, Process, Technology, Data, Governance and Continuous Improvement (2PTDGC) as a Framework to Explore the Cybersecurity Awareness and Process of Australian Medical Practices.

Ferdynandus, F., Prihanto, J. N., & Winarno, W. (2024). Implementing NIST Framework and the People, Process, Technology Approach in Indonesian Financial Services. International Journal of Engineering Continuity, 3(1), 172-182.

Franco, E., Yin, R., & Sankaranarayanan, B. (2024). Building Critical Statewide Cybersecurity Capabilities: The Wisconsin Model. Proceedings of the 25th Annual International Conference on Digital Government Research,

Ganapati, S., & Desouza, K. (2024). Public Value Principles for Secure and Trusted AI. Proceedings of the 25th Annual International Conference on Digital Government Research,

Janowski, T. (2015). Digital government evolution: From transformation to contextualization. In (Vol. 32, pp. 221-236): Elsevier.

Khisro, J. (2025). AI in Digital Government: A Literature Review and Avenues for Future Research.
Knutsen, L. Z., Hannay, J. E., & Riegler, M. A. (2024). Artificial Intelligence in the Public Sector--An Agenda for Responsible Innovation through Learning. Proceedings of the 7th ACM/IEEE International Workshop on Software-intensive Business,
Kumar, A., Shankar, A., Hollebeek, L. D., Behl, A., & Lim, W. M. (2025). Generative artificial intelligence (GenAI) revolution: A deep dive into GenAI adoption. Journal of Business Research, 189, 115160.
Layne, K., & Lee, J. (2001). Developing fully functional E-government: A four stage model. Government Information Quarterly, 18(2), 122-136.
Leblanc, R., & Gillies, J. (2005). Inside the boardroom: How boards really work and the coming revolution in corporate governance. John Wiley & Sons.
Lindgren, I., & Van Veenstra, A. F. (2018, 2018-05-30). Digital government transformation. Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age,
McClelland, D. C., & Boyatzis, R. E. (1980). Opportunities for counselors from the competency assessment movement. The Personnel and Guidance Journal, 58(5), 368-372.
Medaglia, R., Gil-Garcia, J. R., & Pardo, T. A. (2023). Artificial intelligence in government: Taking stock and moving forward. Social Science Computer Review, 41(1), 123-140.
Mergel, I., Dickinson, H., Stenvall, J., & Gasco, M. (2024). Implementing AI in the public sector. Public Management Review, 1-14.
Microsoft. (2025a). AI security for Microsoft 365 Copilot. Microsoft. https://learn.microsoft.com/en-us/copilot/microsoft-365/microsoft-365-copilot-ai-security
Microsoft. (2025b). Data, Privacy, and Security for Microsoft 365 Copilot. https://learn.microsoft.com/en-us/copilot/microsoft-365/microsoft-365-copilot-privacy
Microsoft. (2025c). Microsoft 365 Copilot overview. Microsoft. https://learn.microsoft.com/en-us/copilot/microsoft-365/microsoft-365-copilot-overview
Microsoft. (2025d). Microsoft 365 Copilot requirements. Microsoft. https://learn.microsoft.com/en-us/copilot/microsoft-365/microsoft-365-copilot-requirements
Miles Matthew, B., Huberman, A. M., & Saldana, J. (2020). Qualitative data analysis: A methods sourcebook. In (4th ed.): Sage Publications.
Millan-Vargas, A. O., Sandoval-Almazan, R., & Valle-Cruz, D. (2024). Impact and barriers to AI in the public sector: the case of the State of Mexico. Proceedings of the 25th Annual International Conference on Digital Government Research,
Myers, M. D. (1997). Qualitative Research in Information Systems. MIS Quarterly, 21(2), 241-242. https://doi.org/www.qual.auckland.ac.nz (MISQ Discovery, archival version, June 1997)
NCSC. (2025). Joint Guidance: Principles for security-by-design and -default. NCSC. https://www.ncsc.govt.nz/news/security-by-design
O'Keeffe, J., Buytaert, W., Mijic, A., Brozović, N., & Sinha, R. (2016). The use of semi-structured interviews for the characterisation of farmer irrigation practices. Hydrology and Earth System Sciences, 20(5), 1911-1924.
Pathak, A., & Intratat, C. (2012). Use of semi-structured interviews to investigate teacher perceptions of student collaboration. Malaysian Journal of ELT Research, 8(1), 1.
Peppard, J., & Ward, J. (2004). Beyond strategic information systems: towards an IS capability. The Journal of Strategic Information Systems, 13(2), 167-194.
Peretz-Andersson, E., Lavesson, N., Bifet, A., & Mikalef, P. (2021). AI Transformation in the Public Sector: Ongoing Research. 2021 Swedish Artificial Intelligence Society Workshop (SAIS),
Persson, P., & Zhang, Y. (2025). Openness And Transparency by Design: Crafting an Open Generative AI Platform for the Public Sector. Proceedings of the 58th Hawaii International Conference on System Sciences. https://hdl.handle.net/10125/109067,
Sattlegger, A., & Bharosa, N. (2024). Beyond principles: Embedding ethical AI risks in public sector risk management practice. Proceedings of the 25th Annual International Conference on Digital Government Research,
Straub, V. J., Morgan, D., Bright, J., & Margetts, H. (2023). Artificial intelligence in government: Concepts, standards, and a unified framework. Government Information Quarterly, 40(4), 101881.
Tangi, L., Janssen, M., Benedetti, M., & Noci, G. (2021). Digital government transformation: A structural equation modelling analysis of driving and impeding factors. International Journal of Information Management, 60, 102356.
Toll, D., Lindgren, I., Melin, U., & Madsen, C. Ø. (2019). Artificial Intelligence in Swedish Policies: Values, benefits, considerations and risks. Electronic Government: 18th IFIP WG 8.5 International Conference, EGOV 2019, San Benedetto Del Tronto, Italy, September 2–4, 2019, Proceedings 18,
Ushaka Adie, B., Tate, M., & Valentine, E. Digital leadership in the public sector: a scoping review and outlook. International Review of Public Administration, 1-17. https://doi.org/10.1080/12294659.2024.2323847
Valentine, E. L. (2016). Enterprise technology governance: New information and technology core competencies for boards of directors Queensland University of Technology.
Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial intelligence and the public sector—applications and challenges. International Journal of Public Administration, 42(7), 596-615.

Appendix A: AI risks and mitigation strategies

Below are the codes for AI risks and mitigation strategies analysed from the interview data.

References	Codes	Themes	Dimensions
LCA1, LCA2, LCA5, LDT1, LDT2, PCA1, PCA2, PCA4, PCA6, PDT2	Data Leakage	Data risks	**AI risks**
LCA3, LCA5, PCA3, PDT1	Privacy issues	Privacy risks	
PCA3, PCA7	Data Sovereignty and Jurisdiction issues	Data Sovereignty and Jurisdictional risks	
PCA3	Trusting big AI corporate vendors		
PDT4	Automated decision-making with no human intervention	Technology risks	
LCA2	No Moral Compass		
LCA4	Automation of existing risk (bad behaviour)		
LCA5, PCA4, PDT1	AI Bias		
LCA5, PCA1	AI Hallucination		
PCA1, PCA3	Loss of control when using AI		
PDT6	Trusting AI with big decisions		
PCA2	Human Error and Misuse		
LCA1, LCA5, PCA6, PDT2	Lack of understanding of AI risks	AI strategy and adoption risks	
LCA1, PCA3, PCA5, PCA6, PDT2	Lack of AI knowledge and effective use		
LCA2	Creating our own version (another technical debt)		
LCA3, LDT1, PCA7	Technology hype circle (buzzword)		
LCA4	Fear of adoption and use (risk-averse)		
LCA2, LCA3	No AI Guardrails in place		
LCA4	Lack of AI regulation		
LDT2	Lack of AI security assurance and policies		
PCA3	Lack of Transparency and Consent		
LCA2, PCA2, PCA3	Environmental Sustainability of AI Data Centres	Environmental and sustainability issues	**Other areas of concern**
PCA2	Cost of Adoption of AI		
LCA3, PCA4	AI Funding & Investment	AI Strategy, and policies	
LDT1, PCA1, PDT3, PDT4	AI Policy & Strategy		
PDT4	Human Intervention		**AI risk mitigation strategies**
PDT5	Vendor Management		
LCA1, PCA7	Best practice guidelines	AI best practice standards and guidelines	
LCA1, LCA2, LCA3,	Governance, Guardrails &		

LDT3, PCA7	Standards	
LCA4, LCA5, LDT1, LDT2, PCA1, PCA6, PCA7, PDT1, PDT4, PDT5, PDT6	AI risk assessment & Controls	AI risk assessments and security controls
LDT3, PDT1	Privacy Impact Assessments	
LCA2, LCA4, PCA2, PDT2, PCA5	AI Awareness & Education	AI Awareness & Education