EXPLAIN YOURSELF, AI: WHY CONTESTABILITY IS THE NEXT FRONTIER IN PUBLIC SECTOR ETHICS

RESEARCH INNOVATION COUNCIL ANZ BRIEF | VOL.18

EXECUTIVE SUMMARY

NSW's Research Innovation Council convened senior public sector leaders, technologists, academics, and industry experts under Chatham House rules at the University of Technology Sydney to explore the current and future state of responsible artificial intelligence (AI) in government. The discussion covered ethical governance, misinformation, vendor lock-in, AI literacy, and the integration of AI into complex service delivery environments.

Participants emphasised that AI adoption must not outpace responsible oversight. While AI offers clear efficiencies, especially in targeted service delivery, planning, and internal automation, its use raises complex questions around accountability, explainability, and ethical alignment—particularly in a devolved governance environment. A recurring insight was the need to assess whether AI is objectively better than current processes before deployment, especially from the lens of citizen trust and transparency.

Leaders called for interoperable frameworks, shared procurement assessments, and stronger inter-agency collaboration. They also identified education, contestability mechanisms, and open-source alternatives as key to mitigating vendor lock-in and enabling more agile, ethical AI deployment. The session underlined the importance of empowering public servants—especially non-technical staff—with the literacy and tools needed to evaluate and safely use AI within existing public service values.

FUTURE RESEARCH AND DISCUSSION

AI Literacy at Scale: Participants repeatedly cited the need to accelerate workforce-wide AI understanding. This includes not just tool usage but core concepts such as bias, contestability, and model explainability. As AI becomes embedded in routine tools, ongoing education is essential to prevent unconscious misuse and build ethical reflexes across roles.

Agentic AI and Process Automation: With rapid advancements in AI agents—tools that act autonomously across systems—participants expect major impacts on government workflows. These agents may handle complex interactions such as booking services, processing documents, or executing transactions. While offering efficiency, they raise new concerns around error amplification, traceability, and trust. Future roundtables should explore governance models specific to agentic AI.

Vendor Lock-In and Modular Procurement: Participants expressed growing concern about long-term dependency on dominant AI providers. Future work could explore modular procurement strategies, open source integration pathways, and mutualised assessments of compliance across evolving product lines.

Institutional Contestability and Citizen Trust: Finally, participants called for a greater focus on public trust and contestability—designing systems that citizens understand, question, and feel treated fairly by. As AI is used in licensing, planning, and compliance, the legitimacy of automated decisions will hinge not only on performance, but on transparency and recourse.

By **Patrick Joy** | Head of Research & Analysis | **Public Sector Network**

KEY THEMES AND INSIGHTS

FROM COMPLIANCE TO CONTESTABILITY: REDEFINING RESPONSIBLE AI

Responsible AI is evolving beyond abstract principles to include mechanisms for accountability and contestability. Participants noted that AI-generated outputs must not be accepted at face value—especially where decisions affect citizens. Instead, systems must include explanations that are transparent, faithful to underlying models, and comprehensible to non-expert users. One agency noted their dual-output system that combines conventional AI with generative AI, allowing users to cross-check results against known provenance and judge trustworthiness.

As several participants observed, government must treat AI as an augmentation to human judgment—not a delegation of responsibility. Explainability, they argued, must support contestability. The risk is not just flawed decisions, but loss of public trust when outcomes cannot be challenged or understood.

AI AS A MIRROR: BETTER THAN WHAT?

A central theme was whether AI is demonstrably better than the processes it replaces. Participants urged agencies to rigorously assess the marginal risks and benefits of AI adoption—not just the absolute risks of AI. Automation, they noted, can increase efficiency but may also scale up flawed processes, reduce human discretion, or obscure accountability.

One academic highlighted this with a clear formulation: AI should be judged against the quality and risks of the current human-led process. Efficiency gains must not come at the cost of fairness or transparency. This mindset shift was widely endorsed, reframing responsible AI as a comparative decision rather than an ideological one.

ETHICS IN A DEVOLVED SYSTEM: CENTRAL GUIDANCE VS LOCAL ACCOUNTABILITY

The roundtable surfaced differing views on how ethical guardrails should be implemented. Some participants advocated for stronger centralised oversight, citing the need for standardised procurement reviews and unified assurance mechanisms—particularly to combat vendor lock-in and uneven capabilities across agencies.

Others warned against over-centralisation, arguing that ethical decision-making must remain devolved to ensure context-specific accountability. As one agency lead put it, "the presence of a fancy new tool doesn't eject you from your responsibilities as a delegated decision-maker." Participants ultimately agreed on the need for shared frameworks and transparency, while preserving local accountability for AI use cases.

VENDOR LOCK-IN AND THE OPEN SOURCE HORIZON

Vendor lock-in emerged as a critical structural challenge. With many agencies deeply embedded in single-vendor ecosystems (e.g., Microsoft 365), participants expressed concern about long-term dependency and limited contestability. AI-native platforms are increasingly embedded into productivity tools, complicating governance and procurement oversight.

Open source platforms were highlighted as a counterbalance—offering flexibility, transparency, and community-driven innovation. However, concerns were raised about their lack of standardised compliance with evolving ethical frameworks. Participants called for coordinated investment in open source vetting and support to enable real-world alternatives to proprietary AI infrastructure.

CHALLENGES AND BARRIERS

Skills, Literacy, and Internal Readiness:
A cross-cutting concern was uneven AI literacy across the public sector. While some executives and technology officers are actively shaping AI policy, many frontline staff lack even basic understanding of AI capabilities or risks. One participant likened the gap to early-stage cybersecurity, where years of awareness-raising were required to instil safe practices.

This gap is further complicated by the influx of generative AI into commodity tools—where staff may be using AI without awareness or training. Participants cited examples of staff pasting sensitive information into public models, or unintentionally relying on AI outputs without verification.

Misalignment of Procurement and Risk Practices:
Current procurement models were described as duplicative and poorly suited for AI. Agencies are individually assessing terms of service, vendor claims, and platform risks—often without legal or technical expertise. Many participants endorsed the idea of pre-qualified vendor reviews or shared risk assessments to reduce redundancy and improve security.

A lack of clarity on data residency, platform updates, and embedded AI features (e.g., within Microsoft CoPilot) was also flagged. Some vendors, participants noted, claim AI compliance based on platform features alone, bypassing critical questions around data context, use cases, and deployment risk.

Cross-Agency Collaboration and Interoperability:
AI use cases often transcend agency boundaries, especially in emergency response, infrastructure planning, and customer service. Yet participants described challenges in inter-agency collaboration, citing differences in maturity, policy interpretation, and systems integration.

Despite examples of strong collaboration during crises (e.g., cross-state flood coordination), participants called for more structured forums for sharing use cases, test results, and compliance learnings. Bodies such as the AI Working Group and CIO forums were seen as critical but not yet sufficient for system-wide maturity.

Decision Paralysis in a Crowded Governance Landscape:
Participants described the cumulative burden of overlapping digital reforms—including cybersecurity, privacy, ethics, and records compliance. AI is arriving in an already complex policy environment, leading to uncertainty about what rules apply, who is accountable, and when action can be taken. This cognitive overload contributes to inertia, hesitation, and inconsistent implementation across agencies.

Third-Party AI Use Without Transparency:
Participants reported cases where consultants or vendors embedded generative AI tools into their service delivery—without agencies' knowledge or consent. One example involved a firm including an AI assistant as part of its team roster in a government tender. This creates risks around transparency, data exposure, and quality control, especially where AI output is absorbed into formal government processes without appropriate oversight.

INNOVATIVE IDEAS AND CASE STUDIES

1. IN-HOUSE AI TOOLS TO SAFELY CHANNEL DEMAND

To pre-empt unsafe use of public AI tools, some agencies have developed internal AI platforms that allow staff to experiment within safe parameters. One department launched an in-house assistant powered by vetted models, accessible only through internal networks and trained exclusively on approved datasets.

This approach channels the natural curiosity and productivity demands of staff, while reducing the risk of data leaks or unsanctioned AI use. It also reinforces digital literacy by embedding usage guidelines into the platform itself.

2. CONTESTABILITY IN HIGH-STAKES ENVIRONMENTS

Participants discussed AI deployment in sensitive settings, such as fraud detection or eligibility determinations. One agency described using machine learning to identify vulnerable citizens for proactive debt resolution—not punitive enforcement. Here, AI flagged cases, but human officers made the final call.

Others emphasised traceability and provenance. In legal and investigative contexts, one vendor noted their system allows users to compare AI outputs to known evidence trails, enabling defensibility in court or before commissions.

3. ACADEMIC PARTNERSHIPS FOR RISK ASSESSMENT

Universities such as UTS were praised for co-developing AI ethics frameworks, risk assessment tools, and validation methods. One collaboration mapped misinformation pathways in radicalised groups, identifying intervention points and informing public policy. Another project worked on continuous risk monitoring tied to evolving data and algorithmic behaviour.

Participants endorsed stronger academic partnerships as a means to build public sector capacity, test emerging frameworks, and ground strategy in evidence.

STRATEGIC OUTCOMES AND RECOMMENDATIONS

IMMEDIATE ACTIONS

- **Create Safe Internal AI Environments**: Agencies should offer in-house AI tools for experimentation and routine tasks, with embedded guidance and data protections.
- **Review Procurement Terms Collectively**: Establish inter-agency working groups to standardise AI vendor reviews and address shared concerns (e.g., data use, auto-updates, embedded features).
- **Baseline AI Literacy**: Launch foundational AI training tailored to non-technical staff, paired with regular refreshers like those used in cybersecurity

MEDIUM-TERM GOALS

- **Implement Use Case-Based Governance**: Deploy tiered risk assessment frameworks across AI projects, using examples from the NSW AI Assurance Framework.
- **Build Inter-Agency Collaboration Mechanisms**: Expand existing working groups and CIO forums to focus on shared implementations, auditability, and lessons learned.
- **Strengthen Vendor Accountability**: Develop contractual provisions that require transparency on model updates, data handling, and explainability.

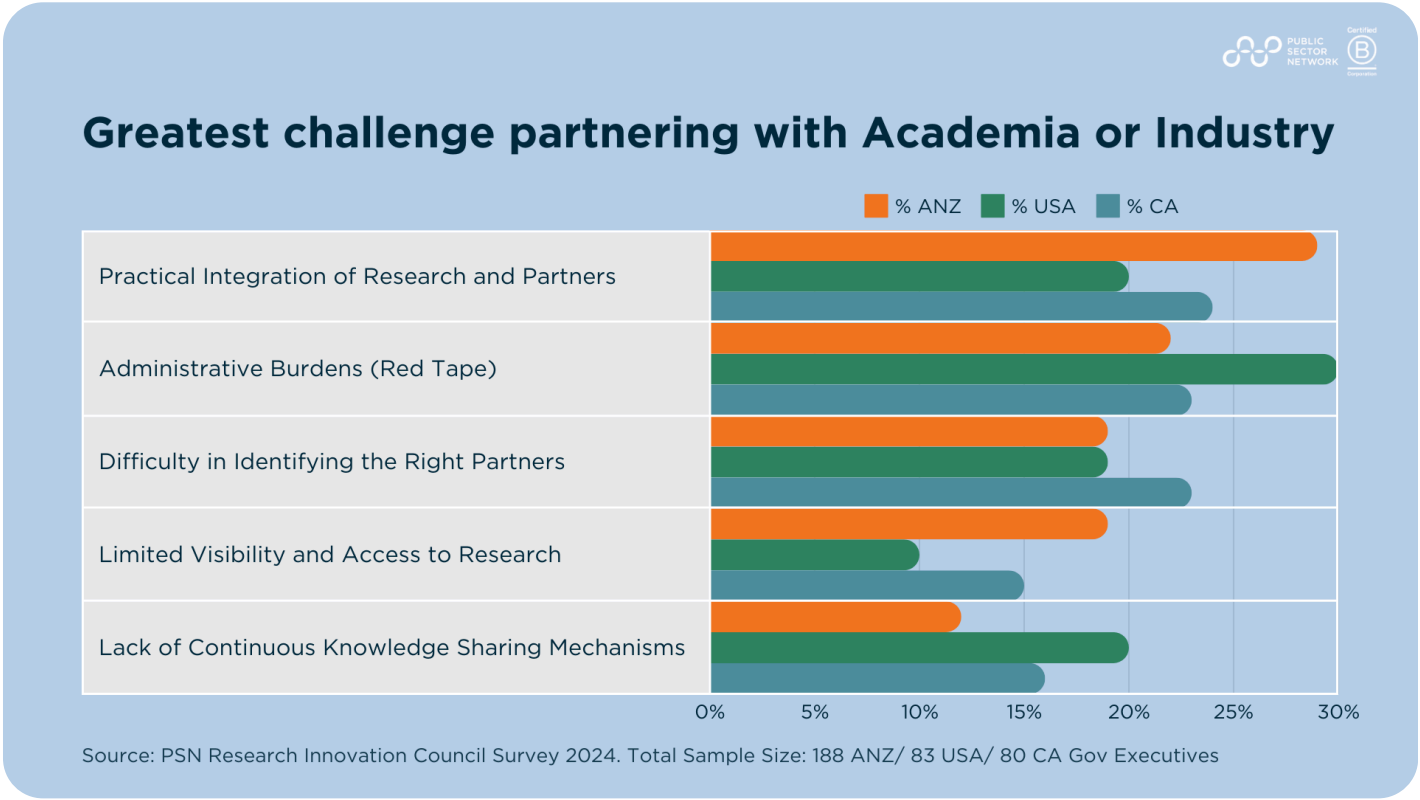
LONG-TERM VISION

- **Embed Contestability into System Design**: Institutionalise the right to challenge AI-driven decisions by requiring explainable outputs, audit logs, and human override mechanisms.
- **Foster Open Source Alternatives**: Invest in community-supported, standards-aligned open source AI tools to reduce reliance on proprietary platforms.
- **Develop Cross-Government Assurance Capabilities**: Establish a central body or shared service that can assess AI systems for security, ethical compliance, and risk management.

ABOUT THE RESEARCH INNOVATION COUNCIL

Connecting government, industry and academia: we've been able to engineer a new program antithetical to the classical red tape, administration, and risk-aversion that impedes innovation.

Our program offers senior executives a unique opportunity to gather monthly for PSN-facilitated roundtable sessions with colleagues, academics, and technologists. We run the program at no cost to government executives or academics, and provide attendees a private online group on our global insights exchange platform, access to research briefs, and expert engagement with PSN analysts and facilitators.



ABOUT PUBLIC SECTOR NETWORK

Public Sector Network is a research company that represents public sector professionals across Australia, Canada, New Zealand, and the USA. It develops roundtables, seminars, and conferences to suit current areas of interest to government agencies and their suppliers.

PSN's growing community spans across federal, state, and local government departments, healthcare, and education, allowing members to share information, access the latest in government innovation, and engage with other like-minded individuals on a secure and closed-door network.

AUSTRALIA / NEW ZEALAND
P +61 2 9057 9070
E INFO@PUBLICSECTORNETWORK.COM.AU

USA
P +1 (647) 969 4509
E HELLO@PUBLICSECTORNETWORK.COM

CANADA
P +1 (647) 459 8904
E CONTACT@PUBLICSECTORNETWORK.CO

Public Sector Network (Australia) Pty Ltd
ABN - 46 617 870 872 20-40
Meagher Street, Chippendale, Sydney NSW 2008, Australia