

Submission to the Joint Select Committee on Artificial Intelligence

Submitted by: Trustethica

Date: 14 September 2026

Prepared by: Harm Ellens and Vivegavalen S Vadi Valu for and on behalf of Trustethica

Disclaimer

This submission has been prepared by Trustethica for the purpose of contributing to the Committee's inquiry. It is provided for policy and technical discussion and does not constitute legal or regulatory advice. References to third-party laws, standards, frameworks or publications are provided for context and do not imply endorsement, affiliation or sponsorship. For feedback and enquiries, contact hello@trustethica.com.

Executive summary

Trustethica welcomes the Joint Select Committee on Artificial Intelligence inquiry and its focus on securing artificial intelligence's economic, social and strategic benefits while managing risks to people, institutions, security and national resilience.

Australia's policy challenge is not simply whether to regulate AI or accelerate its adoption. It is how to make adoption governable in practice. For consequential AI—particularly systems used in regulated, high-consequence or public-facing contexts—governance cannot end with a policy statement, an AI register or a pre-deployment assessment. Governance must continue while systems operate and encounter changing data, users, contexts, capabilities, suppliers and incentives.

Trustethica submits that Australia should adopt an assurance-led, risk-proportionate and technology-neutral approach. Consequential AI systems should be connected to an authorised purpose, a named accountable owner, defined risk thresholds, ongoing monitoring, controlled exception handling and auditable evidence.

This approach would provide practical adoption infrastructure: it would give organisations a clearer route to deploy beneficial AI with confidence; strengthen transparency and accountability in public administration and regulated sectors; support meaningful Australian control over data, governance decisions and critical dependencies; and address safety, privacy, consumer, discrimination, cybersecurity, fraud, intellectual-property and national-security risks in a coherent operational model.

Trustethica recommends that the Committee support a nationally consistent AI assurance framework, a common accountability record for consequential systems, runtime-governance expectations for high-consequence and agentic AI, controlled pilots with published implementation learning, and an Australian governance and assurance profile aligned to relevant international standards.

From AI principles to operational accountability

The inquiry's objectives, i.e., productivity, national capability, public-service adoption, workforce impacts, legal adequacy, data sovereignty, safety, cybersecurity and national security all meet at the point of organisational deployment. An AI system becomes consequential when it influences a material decision; changes access to an important service; affects an individual or community; produces public-facing information at scale; handles sensitive information; enables a critical process; or has meaningful security implications.

In these settings, the practical question is not only whether a model is capable or whether an organisation has adopted responsible-AI principles. It is whether the organisation can demonstrate what the system is authorised to do and not do; whose purpose it serves; who is accountable for it; which data, models, tools, suppliers and interfaces form part of the operational system; which risks and controls apply; and how incidents, complaints, exceptions, overrides and residual-risk decisions are managed.

Principles-based policy remains necessary, but principles alone do not allocate responsibility, detect changed conditions, preserve evidence of management action or demonstrate that controls are operating. Australia should complement principles and pre-deployment assessment with an operational governance model that is observable, reviewable and proportionate throughout an AI system's lifecycle.

Trustethica's proposed foundation is purpose-anchored contextual AI governance. An AI system should operate within an explicitly authorised purpose and context. For material use cases, this purpose should be linked to a named accountable executive and operational owner, affected stakeholders, defined boundaries, applicable obligations, risk tolerances and escalation arrangements.

For consequential systems, organisations should maintain a structured accountability record. Proportionate to the risk, this record should identify the authorised purpose and limitations; intended users and affected persons; accountable owners; the system's role and degree of human authority; material model, data, tool, vendor and infrastructure dependencies; relevant legal and policy obligations; testing and assurance evidence; monitoring indicators and escalation thresholds; material incidents and changes; residual-risk acceptances; and review or decommissioning criteria.

This need not create a new paperwork burden. Properly designed, a common record enables evidence to be reused for internal governance, procurement, supplier assurance, audit, regulatory engagement and incident response.

Why operational assurance matters

Productive and inclusive adoption

Responsible governance is adoption infrastructure, not merely a compliance cost. Organisations often delay valuable AI deployment because they cannot determine who is accountable, what controls are required, whether suppliers are acceptable, how privacy and security obligations apply, or what evidence will be needed if problems arise.

A common assurance model reduces this uncertainty. It gives boards and executives a repeatable basis for approving use cases, gives operational teams workable guardrails, and gives small and medium enterprises a clearer route from experimentation to trusted deployment. Obligations should scale with the likely severity, reach and reversibility of harm, rather than impose the same burden on low-risk internal productivity tools and systems affecting legal, financial, health, safety or other significant interests.

Sovereign capability and public accountability

Australia's AI capability depends on more than access to frontier models. It includes domestic skills, trusted deployment practices, resilient supply chains, local commercial capability and the capacity to retain meaningful control over sensitive data and critical decisions.

Governance infrastructure is therefore part of sovereign capability. Australian organisations should be able to understand and evidence where their data is used, who can access it, which systems and suppliers influence an AI outcome, which obligations apply, and which accountable person can halt or constrain use. For sensitive, regulated or strategically significant deployments, policy should

encourage patterns that preserve customer control, support data-residency and isolation requirements, minimise unnecessary data transit, retain auditable logs and make material supplier dependencies transparent.

Commonwealth agencies can use AI to improve accessibility, service quality, timeliness, administrative efficiency and staff capability. However, systems that affect rights, entitlements or access to services require a particularly high standard of transparency, contestability and accountability. Government should demonstrate clearly stated public purposes, meaningful human responsibility, accessible pathways for challenge and redress, documented risk decisions, independent review where warranted, and public reporting that is intelligible without exposing security-sensitive information.

Safety, security and agentic AI

AI can improve cyber defence, anomaly detection, service triage and threat analysis. It can also enable more convincing fraud, scalable social engineering, deepfakes, automated vulnerability discovery, misuse of personal information and novel forms of foreign interference.

For material deployments, AI risk management should connect with established security and resilience practices. This includes recording material dependencies; applying access controls; protecting logs and assurance evidence; identifying abuse and misuse scenarios; monitoring for policy violations; establishing incident escalation arrangements; and conducting structured post-incident review.

Runtime governance is particularly important where systems can act autonomously or semi-autonomously, interact directly with customers or citizens, affect significant interests, or operate in high-fraud, cyber, national-security or critical-infrastructure environments. Organisations should be able to monitor systems against authorised purpose and risk thresholds, detect material deviation, route issues to accountable people, preserve evidence and make documented decisions about intervention or continued operation.

For agentic AI, Australia should set stronger expectations around explicit authority boundaries, tool access, approval gates, segregation of duties, action logging, revocation mechanisms, fail-safe behaviour and review of unexpected or high-impact actions.

Recommendations

Trustethica respectfully recommends that the Committee support the following measures.

1. **Establish a nationally consistent, risk-proportionate AI assurance framework.** The Australian Government should develop a technology-neutral framework that differentiates low-risk AI use from consequential and high-risk use. It should establish baseline expectations for accountability, recordkeeping, risk assessment and incident management, with enhanced requirements where potential harms are substantial, scalable or difficult to reverse.
2. **Require purpose, ownership and evidence for consequential AI.** Organisations operating consequential AI should maintain a proportionate accountability record covering authorised purpose and boundaries, accountable ownership, affected persons, material dependencies,

applicable obligations, risk classification, controls, monitoring, significant changes and residual-risk decisions. Material expansion of scope, autonomy, audience, data use or decision authority should trigger reassessment and appropriate approval.

3. **Set minimum runtime-governance expectations for high-consequence and agentic AI.** Significant systems should be subject to ongoing monitoring against defined indicators and thresholds, traceable escalation, controlled exception handling, action and decision logging, human-intervention mechanisms and preservation of evidence. Agentic systems should additionally be subject to explicit controls over authority, tool access, approvals, traceability, segregation of duties, revocation and safe failure.
4. **Fund controlled pilots and publish implementation learning.** Government should support time-bounded pilots in Commonwealth services and regulated or nationally significant sectors. Evaluation should address not only technical performance and productivity, but also control effectiveness, human oversight, impacts on affected people, incident handling, implementation cost, interoperability and assurance-evidence quality. Where safe and appropriate, government should publish de-identified lessons, reusable control patterns and procurement guidance.
5. **Create an Australian AI governance and assurance profile aligned to international standards.** Government, regulators, standards bodies, industry, civil society and technical experts should develop an Australian profile compatible with relevant international standards, including ISO/IEC 42001 for AI management systems, while reflecting Australian law, public-interest expectations and sectoral requirements. The profile should provide a simple baseline for SMEs and low-risk adoption; enhanced controls for systems affecting significant interests or operating at scale; stronger expectations for agentic, public-sector, critical-infrastructure and security-sensitive systems; and reusable evidence templates for procurement, suppliers, internal audit and regulatory engagement.

Conclusion

Australia has an opportunity to make trusted AI adoption a national strength. This requires more than aspirational principles and more than a choice between innovation and regulation. It requires practical governance infrastructure through which organisations can deploy beneficial AI with clear purpose, accountable ownership, proportionate controls and demonstrable evidence.

Trustethica submits that the Committee should support a national approach built around purpose-anchored, risk-proportionate and operationally testable governance. Such an approach would support productivity and innovation, strengthen sovereign capability, protect people and institutions, improve public confidence, and help Australia respond to safety, cyber and national-security risks.

In concise terms, consequential AI should not merely be approved before deployment; it should remain accountable throughout its operation, with a direct and continuous line of sight between runtime AI activity, authorised purpose, AI-specific risks, and defined risk thresholds.

About Trustethica

Trustethica provides purpose-anchored contextual AI governance infrastructure for regulated and high-accountability enterprise environments. The platform governs AI at the contextual use-case level, where business purpose, operational behaviour, and enterprise liability intersect. Its approach brings together AI governance, enterprise risk management, technology law, privacy, data governance, cybersecurity and assurance in an operational model for accountable AI deployment.