



Ownership, Not Reach

REFERENCE	AOS-AAH-2026-001
VERSION	1.0
DATE	13 September 2026
CLASSIFICATION	Released
DIRECTOR	Brunei Al-Bijwaie

Ownership, Not Reach

Version 1.0. This assessment reads four publications of the Australian Signals Directorate on artificial intelligence — *Agentic AI Harnesses* — *The layer above the model*, *Careful adoption of agentic AI services*, *Frontier AI cyber threat considerations for boards of directors*, and *Opportunities for AI in cyber defence* — whole, from eight pinned captures, as at the retrieval dates in the references.

All four reach supply. *Opportunities for AI in cyber defence* asks vendors to demonstrate model provenance and lineage, to provide a software and AI bill of materials meeting the minimum elements, to state what the capability depends on in third-party models, services and data sources, and to say which jurisdictions process or store customer data; it also requires organisations to map their training and fine-tuning data sources to assess provenance. *Frontier AI cyber threat considerations for boards of directors* asks boards whether they rely on vendors and service providers without an understanding of their foreign ownership, control or influence, and whether they have visibility of third and fourth-party suppliers. *Agentic AI Harnesses* requires the reader to select a trusted model and assess its provenance. *Careful adoption of agentic AI services* requires that third-party components originate from trusted sources, that a trusted registry be kept, and that bill-of-materials minimum elements be referenced when procuring.

None of the four asks who can lawfully require production of a model, its weights, its logs, or the data it processes. The words *compel* and *lawful* occur in none of them.

In three of the four, the supply question sits in a register other than the one that tells the reader what to do. *Agentic AI Harnesses* requires provenance be assessed in its good practice and asks nothing about it in the seven questions it puts to executives. *Careful adoption* requires trusted-source origination in its best practices and puts no questions at all. *Frontier AI boards* asks about foreign ownership in its threshold questions, and the words *foreign*, *ownership* and *influence* do not appear after line 123, where its four priority registers and its practical actions begin — including in the priority that instructs boards to oversee the deployment of AI models into software development, security assessment and security monitoring, where the whole supply question is carried by the word *suitable*, three times, and the document does not define it.

This assessment establishes what these four instruments ask and require. It does not establish what the Directorate asks elsewhere, does not argue that any omission was deliberate, and does not claim the four form a pattern.

Scope

This assessment reads four publications of the Australian Signals Directorate on artificial intelligence, whole, and establishes what each asks and requires about who supplies, owns, hosts or can lawfully reach the models and systems it addresses.

The four are *Agentic AI Harnesses — The layer above the model*, twelve pages, first published September 2026; *Careful adoption of agentic AI services*, twenty-nine pages, co-authored by the Directorate's Australian Cyber Security Centre with the United States Cybersecurity and Infrastructure Security Agency and National Security Agency, the Canadian Centre for Cyber Security, the New Zealand National Cyber Security Centre and the United Kingdom National Cyber Security Centre; *Frontier AI cyber threat considerations for boards of directors*, seven pages, co-published with the Australian Institute of Company Directors, August 2026; and *Opportunities for AI in cyber defence*, twenty-four pages, co-sealed with the Canadian, New Zealand and United Kingdom centres. Eight captures are pinned, two for each instrument, each hashed. Where this case retrieved a capture it was fetched twice under the same agent within the minute and hashed on both fetches; the four it did not retrieve are pinned by the registers named in the references, and neither of those registers records a second fetch. The retrieval dates and hashes are in the references.

The population's basis

The population is **the subject, the instrument the subject incorporates by reference, and the two Directorate publications on artificial intelligence already pinned in this corpus.**

It is not “guidance addressed to executives”, and the reason is worth stating because it decides what belongs. *Agentic AI Harnesses* addresses “executive decision-makers, chief information security officers and IT leaders”. *Frontier AI cyber threat considerations for boards of directors* addresses boards. *Opportunities for AI in cyber defence* addresses “cyber defenders, such as Chief Information Security Officers and senior security leaders”. But *Careful adoption of agentic AI services* **states no executive audience at all**: its practice sections are addressed to “agentic AI developers”, and to “vendors and operators”. A population defined by audience would exclude it.

It is in the population because the subject puts it there. *Agentic AI Harnesses* names it in its own *Purpose*, states that “The harness publication complements that guidance”, links it five times from the subject's landing page, and summarises its recommendations. Its audience is named wherever a finding about it is reported, so that no finding about guidance to developers is mistaken for a finding about guidance to executives.

The distinction the assessment turns on

Three questions are near neighbours and are not the same question.

Ownership asks who holds a supplier's shares, and who controls or influences it. **Location** asks where data is processed or stored. **Compellability** asks who can lawfully require production of a thing — the model, its weights, its logs, or the data it processes — and on whose application.

An answer to the first two does not settle the third. A supplier wholly owned within one jurisdiction may hold data in another. A supplier holding data within one jurisdiction may be subject to the orders of another. **Neither ownership nor location establishes who can lawfully require production**, and this assessment does not treat either as though it did.

What this assessment establishes

That all four instruments reach supply, that three of them impose a requirement about it, that one asks about ownership and one about location, and that **none asks who can lawfully require production**. The terms *compel* and *lawful* occur in none of the four.

1 Agentic AI Harnesses

Agentic AI Harnesses — The layer above the model is a twelve-page publication of the Australian Signals Directorate, first published September 2026 on the document's own face. Its stated purpose is to explain "the concept of the Agentic AI harness (referred to hereafter as the harness), the software layer that enables a large language model (LLM) to interact with data, tools and systems to perform agentic tasks", and to examine "how harness design influences the security, governance, reliability and operational risks of agentic AI systems". Its stated audience is "organisations adopting or considering agentic AI", and it "is written for executive decision-makers, chief information security officers and IT leaders". Two captures are pinned: the landing page as served, AAH-1, and the publisher's PDF, AAH-2.

1.1 The questions the document puts to executives

The document carries a section headed *Governance questions for executives*. It is prefaced: "Directors and senior executives do not need to understand every implementation detail of an agentic AI system. However, they should seek assurance that the harness, its integrations and its governance controls have been designed and implemented appropriately." Seven questions follow, reproduced here as printed.

What business outcome is this agentic AI system intended to achieve, and why is an agentic approach required?

What data, systems and tools can the agent access, and are least-privilege principles being applied?

What actions can the agent perform, and which actions require human approval?

How are prompt injection, data poisoning and other AI-specific attacks being mitigated?

Can all significant decisions, tool invocations and actions be monitored and audited?

How will we know whether the system is delivering value, operating safely and remaining within acceptable risk tolerances?

If the harness were compromised, misconfigured or manipulated, what is the worst outcome and what controls would prevent or limit that outcome?

1.2 Classification of the seven

Each question is classified by the direction of the relationship it asks about, on the words of the question alone. The test is stated before it is applied so a reader may disagree with a classification on the record.

Inward-reach asks whether a party outside the organisation can obtain, access or compel the model, its weights, its logs, or the data it processes. **Outward-reach** asks what the system itself can reach: what data, tools or systems it touches, or what actions it takes. **Internal traceability** asks whether the organisation can see, record, reconstruct or evaluate what happened inside its own deployment. **Purpose** asks why the system exists. **Threat mitigation** asks how an identified attack class is defended against. **Supply or provenance** asks who supplies, owns, controls or influences the model or its vendor, or where processing occurs. A question is classified on what it asks, not on what answering it might incidentally require.

On that test: questions two and three are outward-reach; questions five and six are internal traceability; questions four and seven are threat mitigation; question one is purpose. The counts are **two outward-reach, two internal traceability, two threat mitigation, one purpose, zero inward-reach, zero supply or provenance**.

The nearest question to inward-reach is the second, and it asks the opposite direction: what the agent can reach, not who can reach the agent.

1.3 What the document says about supply, outside the questions

Six passages outside the question set bear on supply, origin or provenance. They were established by a case-insensitive substring search over the full extracted text of AAH-2, twenty-one terms, and are set out with the register each sits in.

REF-1, narrative. At lines 107 to 110 the document states that “In some cases, the harness is developed by the organisation; in others it is supplied as part of a commercial product or service. Regardless of its origin, organisations should understand the harness’s capabilities, permissions, integration points and security controls, and assess the risks associated with its use.” Origin is raised and then set aside. None of the four things the reader is directed to understand concerns who supplied the harness.

REF-2, good practice. At lines 283 to 285, in the section headed *Good practice*: “Select a trusted model appropriate to the task and assess its provenance, limitations, security characteristics and behaviour. Do not rely on model safety controls instead of harness-enforced controls.” This is the document’s requirement about provenance. It is a bullet in a practice list, not a question, and provenance is not defined, no test for it is stated, and no party is named to perform the assessment.

REF-3, component description. At line 157 the component table describes the model interface as providing “the interface between the harness and organisation-approved LLM services”. The approval is presumed. The document does not state who approves, on what basis, or against what.

REF-4, good practice. At lines 273 to 277, in a cost-control bullet: agentic sessions cost more than a single exchange “whether billed as tokens by a hosted provider or consumed as computing resources on self hosted infrastructure”. This is the document’s only acknowledgement that a third party may run the model, and the

distinction is drawn for cost monitoring and anomaly detection.

REF-5, component description. At lines 176 to 177 the component table carries the row “Update and supply chain path — Introduces new functionality and potential supply-chain risk.” The table’s closing sentence states that these components “are the organisation’s configuration surface”.

REF-6, narrative. At line 244, within a summary of the companion publication’s recommendations, “supply-chain assurance” appears as one item in a list of eight practices attributed to that document.

1.4 The join

The document knows the harness may be supplied by another party, says so, requires the model’s provenance be assessed as good practice, and asks the executive nothing about it: none of the six passages appears in the seven governance questions.

1.5 The negative

The terms *jurisdic*, *sovereign*, *foreign*, *compel*, *lawful*, *offshore*, *extraterritorial* and *overseas* occur zero times in the document. The test was a case-insensitive substring search over the full extracted text of AAH-2, and it was performed twice against independent extractions — the text file and the PDF read directly — returning zero on both.

The limit, at full width: this is a substring test over eight stems. It establishes that those stems do not occur. It does not establish that no passage carries the concept in words none of the eight matches. That question is addressed by the concept test reported at section 5.4, which graded this document NO_MATCH at phrase set 1.0.1 with no concept group producing any hit. That test carries its own limit, recorded under Limitations.

2 Careful adoption of agentic AI services

Careful adoption of agentic AI services is a twenty-nine-page publication co-authored by the Australian Signals Directorate’s Australian Cyber Security Centre with the United States Cybersecurity and Infrastructure Security Agency and National Security Agency, the Canadian Centre for Cyber Security, the New Zealand National Cyber Security Centre and the United Kingdom National Cyber Security Centre. The document refers to these bodies throughout as “the authoring agencies”. Two captures are pinned: the landing page as served, AAH-3, and the publisher’s PDF, AAH-4.

It is in this assessment’s population because the subject incorporates it by reference. *Agentic AI Harnesses* names it in its own *Purpose*, states that “The harness publication complements that guidance”, links it five times from the harness landing page, and summarises its recommendations at lines 241 to 247. **Its audience is not executives.** The document states no executive audience, and its practice sections carry their own audience lines: the first reads “Audience: This section is most relevant to agentic AI developers. Vendors and operators may...”, and two later sections are addressed to “agentic AI vendors and operators”. No finding in this section is a finding about guidance addressed to executives.

2.1 Structure

The document's table of contents gives its shape. An introduction with scope and audience; *Broader agentic AI security considerations*; **five risk categories** — privilege, design and configuration, behaviour, structural, and accountability; *Best practices for securing agentic AI systems* across four lifecycle parts, designing, developing, deploying and operating; *Defend against future risks*; a conclusion; and Appendix A, *Cyber security prerequisites before implementation of AI agents*.

It carries no question set. There is no section of questions for executives or for any other reader. Risks are described in one register and requirements are imposed in another, and that division matters to what follows.

2.2 Two registers on third parties

The document addresses third parties in two places, under headings that differ by one word.

Third-party components, at page 11, sits inside the *Structural risks* category. It states that “Agentic AI systems can introduce structural risk through interactions with third-party components, including tools and other agents”, and enumerates four ways risk arises: “malicious actors engaging in tool or agent ‘squatting’ by publishing malicious tools or agents with legitimate or similar names”; “developers introducing vulnerabilities through configuration errors or insecure third-party components”; “users or systems submitting requests to the incorrect place”; and “tools and agents dynamically loading new packages, increasing exposure to untrusted code”. It closes: “Compromised third-party components can be very difficult to detect due to the limited transparency of agentic AI systems.” Every risk named is a compromise risk, and the qualifying condition given — “depending on the trust level and permissions of the component” — is a configuration property the deploying organisation sets. **This register requires nothing.**

Manage third-party components, at page 17, sits inside the best practices. It states that “Verification and management of third-party components of agentic applications can reduce the added risk”, and under *Recommended best practices* it requires:

Verify all external third-party components originate from trusted sources and are up to date before inclusion in agentic AI systems

Maintain a trusted registry of third-party components

Reference CISA's A Shared Vision of Software Bill of Materials (SBOM) for Cybersecurity and 2025 Minimum Elements for a Software Bill of Materials (SBOM) when procuring agentic AI systems

Appendix A adds a further prerequisite: “Use supply chain risk management practices for third-party dependencies.”

The distance between those two headings is one word, and it is the distance between a register that describes a risk and a register that imposes a requirement. A reading confined to the first would conclude that the document treats third parties only as an attack surface. The second shows that it does not.

2.3 What the requirement does not define

The requirement is that components “originate from trusted sources”. **The document defines *trusted source* nowhere.** No test for trust is stated. No party is named to assess it. Nothing in either register asks or requires anything about who owns or controls a supplier, where a component or service is hosted, which jurisdiction it sits in, or who can compel its producer.

2.4 Accountability risks

The fifth risk category is *Accountability risks*, at page 12. Its lead states that “Agentic system architecture can obscure what caused a particular action, making accountability hard to trace.” Three named items follow, and a scenario example.

Actions and processes states that agent actions and decision-making processes “can be opaque, making agentic AI systems difficult to understand, monitor and audit”, that agents “may initiate secondary tasks, spawn sub-agents, or follow extended delegation chains in ways that are not always visible to operators”, and that “comprehensive logging of agentic AI systems can be difficult, as long reasoning chains and large amounts of contextual data lead to substantial log sizes”, “making extraction of meaningful signals from logs even more challenging”.

Its **scenario example** describes multiple agents approving payments where an erroneous outcome occurs, and states that “it becomes difficult to determine which component or design choice caused the error. Further, fragmented logs, opaque agent reasoning and emergent interactions obscure the decision path, making it hard to explain the outcome, assign responsibility, or demonstrate compliance.”

Accuracy concerns hallucination and reliability: models “may incorrectly interpolate or ‘hallucinate’ plausible-sounding responses when their internal knowledge is insufficient”, and “The system often does not clearly identify this in its output, reducing overall accuracy and reliability.”

Visibility states: “Maintain visibility on status and behaviour when integrating any new system... Agentic system processes can outpace human monitoring capability, possibly leading to unnoticed malicious behaviour, uncaught hallucinations or other issues. While tools perform many actions for an agentic system, they may operate outside of the system’s monitoring boundary, making it difficult to account for tool actions. Additionally, malicious or compromised agents could use tools as a stealthy way to exfiltrate data. Malfunctioning tools could also unintentionally leak data, which may go unnoticed.”

Every item in the category concerns the organisation’s own sight of its own system, and one item names data exfiltration as a consequence of that failure. That is stated as it reads. It is not classified as internal or external, because the item’s subject throughout is the monitoring boundary and what goes unnoticed, while the outcome it names is data leaving, and deciding between those would report a judgment as a measurement.

The category was established by a case-insensitive substring search over the full extracted text of AAH-4, fourteen terms, thirty matches, of which seven sit inside the category and twenty-three outside it. The wider search found a **second accountability treatment** the category does not contain: an *Accountability* best-practice section at page 17, requiring artefact logging by default, “unified audit logs for all inter-agent

interactions”, interpretability tools, and information referencing that shows “where key aspects of their response originated from”. It is also entirely concerned with the organisation’s own sight of its own system.

2.5 The nearest outward legal question in the population

At line 725, under a best-practice heading *Governance*, the document requires: “Define legal accountability and risk ownership for agentic AI systems in policies.”

This is the closest any instrument in this assessment’s population comes to an outward legal question, and it is a drafting instruction rather than an enquiry. It directs the reader to define legal accountability in their own policies. It does not ask who holds the organisation to account, under what law, or on whose application. That is stated rather than classified.

2.6 The negative

No concept group of phrase set 1.0.1 produced any hit of a foreign dimension across the document’s twenty-nine pages. The limit, at full width: that is a concept test over forty-one phrases in five groups under a proximity rule, and the phrase set has one demonstrated gap, recorded under Limitations. It establishes that these concepts, in these phrasings and under that rule, do not occur. It does not establish that no passage carries the concept in words the set does not hold.

3 Frontier AI cyber threat considerations for boards of directors

Frontier AI cyber threat considerations for boards of directors is a seven-page publication of the Australian Signals Directorate, co-published with the Australian Institute of Company Directors, first published August 2026 on the document’s own face. Its audience is boards of directors. Two captures are pinned: the publication page as served, AT-ASD-1, and the publisher’s PDF, AT-ASD-2, both retrieved 5 August 2026.

3.1 Foreign ownership, control or influence, twice

The document raises foreign ownership, control or influence in two places.

In narrative, at lines 51 to 52:

Boards should also be conscious that the implications of frontier AI models extend beyond their effect on cyber security operations. For example, boards should consider cyber supply chain risks associated with their reliance on particular AI vendors and the foreign ownership, control or influence risks this may pose.

As a threshold question, at lines 81 to 82, a sub-bullet under the question “Where could minor weaknesses in our systems and cyber supply chain be combined into a major cyber security incident?”:

Are we relying on vendors and service providers without sufficient governance and oversight, including an understanding of their foreign ownership, control or influence?

The same question set carries a further sub-bullet at line 84: “Do we have visibility of the security and resilience posture of third and fourth-party suppliers within our cyber supply chain?” The question reaches

third and fourth-party suppliers by its own terms.

3.2 The boundary

The document's structure divides at a stated point. After the threshold questions it carries a section headed *What should boards focus on?*, whose introduction at lines 118 to 120 reads:

To give effect to a secure and resilient posture that is capable of responding to frontier AI threats, boards should oversee and challenge management on how their organisations are implementing the following strategies, spanning immediate priorities through to longer-term horizons.

The strategies begin at line 123, under the heading Immediate priorities. Four priority registers follow — Immediate, Short-term, Medium-term and Longer-term priorities — and then a fifth register, *What practical actions can boards oversee?*, at line 166, which runs to line 271. Those five registers are where the document tells boards what to do. The boundary is line 123 and it is stated here rather than characterised.

3.3 The counts after the boundary

Over lines 123 to 315, established by case-insensitive substring search over the full extracted text of AT-ASD-2:

term	lines 1–122	lines 123–315
foreign	2	0
ownership	2	0
influence	2	0
vendor	2	0
supplier	1	0
fourth-party	1	0
supply chain	4	0
provenance	0	0
procure	0	0
hosted	0	0
jurisdic	0	0
compel	0	0
lawful	0	0

Two before, zero after.

The term `control` is reported separately because it occurs on both sides of the boundary and its sense

differs. It occurs three times before line 123 and six times after. Every occurrence after the boundary is the security-control sense: “compensating controls” at line 135, “control access to systems” at 139, “controllable” at 156, “access controls” at 206, “authentication controls” at 215, and “enforce controls consistently” at 260. The single occurrence of `external` after the boundary, at line 213, is “from untrusted, external or high-risk locations”, a network location.

3.4 The medium-term priority

One of the five registers instructs boards on the adoption of AI. The `Medium-term priorities` register, at lines 155 to 156, reads in full:

Adopt AI for cyber defence purposes: Artificial intelligence is used for cyber defence and is secure, controllable, human-supervised and used in an ethical and accountable manner.

That priority becomes three practical actions in the fifth register, at lines 240 to 253, under the same heading:

Deploying suitable AI models for augmenting software development activities, such as identifying and remediating vulnerabilities and weaknesses for software before its release and periodically throughout its life.

Deploying suitable AI models for augmenting security assessment activities, such as performing vulnerability scanning and vulnerability assessments.

Deploying suitable AI models for augmenting security monitoring activities, such as identifying and triaging cyber security events.

3.5 What qualifies the models

Lines 240 to 254 were searched for `supplies`, `supplied`, `supplier`, `supply`, `vendor`, `provider`, `owns`, `owner`, `ownership`, `host`, `hosted`, `hosting`, `foreign`, `jurisdic`, `provenance`, `origin`, `procure`, `acquir`, `trusted`, `reach`, `compel`, `lawful`, `third party` and `third-party`. The only terms matching are `model`, `suitable`, and `assess` — the last in “security assessment activities” and “vulnerability assessments”, which is the activity being augmented and not an assessment of the model.

The supply question in the instruction to deploy AI models into software development, security assessment and security monitoring is carried entirely by the word `suitable`, three times, and the document does not define it.

3.6 What this section establishes, and its limit

The document raises foreign ownership, control or influence as a consideration and as a question, and the term does not appear in any register that tells boards what to do, including the register that instructs them to deploy AI models inside the organisation’s own security function.

The limit, at full width: this is a substring test over named terms across a stated boundary. It establishes that those terms do not occur after line 123. It does not establish that no passage in those five registers carries the concept in words none of the terms matches. The concept test that addresses that question is reported at

section 5.4. It graded this document NO_MATCH at phrase set 1.0.1 with no concept group producing any hit, and the document has since been read end to end for synonyms the phrase set does not hold and signed silent on that basis, with the signoff artefact recorded at section 5.4 and its stated limits under Limitations.

4 Opportunities for AI in cyber defence

Opportunities for AI in cyber defence is a twenty-four-page publication of the Australian Signals Directorate, co-sealed with the Canadian Centre for Cyber Security, the New Zealand National Cyber Security Centre and the United Kingdom National Cyber Security Centre. Its stated audience is at lines 53 to 57: the guidance “supports cyber defenders, such as Chief Information Security Officers and senior security leaders who are responsible for cyber security strategy, operations and risk outcomes”, and it “assumes a foundational understanding of cyber security concepts”. Two captures are pinned: the publisher’s PDF, A-GUID-1, and the guidance page as served, A-GUID-2.

This is the instrument in the population that asks the most about supply, and this section reports what it asks before reporting what it does not.

4.1 Appendix A: questions for AI vendors

The document carries two appendices of questions. Appendix A is introduced at lines 604 to 607: it “provides a set of questions to help organisations assess the security, reliability and maturity of AI capabilities offered by vendors. These questions are designed to distinguish engineered, production-ready solutions from those relying on automation, assumptions, or emerging features.”

Five of its questions bear on supply or jurisdiction. Each is quoted as printed, with its guidance note where one follows.

Model provenance, line 697: “Can the vendor demonstrate clear model provenance and lineage throughout the AI lifecycle?” Note: “This supports assurance, traceability and incident investigation.”

Bills of materials, lines 694 to 696: “Does the vendor provide a machine readable SBOM and AIBOM for AI that meets the minimum elements?” Note: “This supports assessment of AI supply chain transparency and dependency risk.”

Third-party dependencies, lines 680 to 682: “What dependencies does the AI capability have on third-party models, services, or data sources?” Note: “This helps assess supply chain risk, resilience and exposure to upstream changes.”

Portability, lines 691 to 693: “How portable are the AI capabilities if the organisation changes vendors or deployment models?” Note: “This helps identify lock-in risks, particularly where detections or workflows are proprietary.”

Jurisdiction, line 663, under the heading *Data protection and sovereignty*: “Which jurisdictions process or store customer data, including through third-party AI services?” Note: “This is critical for organisations with cross border or regulated data constraints.”

These are questions for AI vendors, in an appendix, and the appendix says so. That is stated because it bears on the level at which the document puts the question: it is put to a supplier about a product, not to an executive about the organisation's exposure.

4.2 A requirement in the body

The question set is not the document's only treatment. In the *Identify* function section, at line 229, under an "Organisations should" list, the document **requires**: "map AI training and fine-tuning data sources to assess provenance, retention and poisoning". At line 228 the same list names "hidden risks, and supply chain exposures".

4.3 AI Secure by Demand

At lines 552 to 558 the document carries a section headed *AI Secure by Demand*:

Organisations should select, implement and operate AI systems with security built in from the outset, not added as an afterthought. This is the Secure by Demand principle, in support of the Secure by Design foundations, and it is relevant to the entire AI supply chain. It means that AI suppliers (and intermediaries) have a responsibility to develop products that are Secure by Design. Secure by Demand means that the customers (and intermediaries) should hold the suppliers to account for that responsibility by only purchasing Secure by Design products.

This is a purchasing instruction. Its object — "Secure by Design products" — is defined in the document only by reference to other documents, which it lists as ASD's *Secure by Design foundations*, the ISM's *Guidelines for software development*, and *Choosing secure and verifiable technologies*.

4.4 The three sovereignty mentions, and what attaches to each

The word **sovereign** occurs three times, and what follows from each differs.

A mitigation bullet, lines 516 to 518, in the *Supply chain* section under "Organisations should": "understand how organisational data is accessed, used and handled across the AI supply chain to ensure alignment with data sovereignty requirements and to determine which types of AI solutions are most suitable." A requirement attaches: to understand, and to ensure alignment. **What the data sovereignty requirements consist of is not stated, no test for alignment is given, and no party is named to assess it.**

A heading, line 657: *Data protection and sovereignty*. Nothing attaches to a heading; it organises the three questions beneath it.

A guidance note, lines 658 to 660. The question is "How is sensitive data handled, stored and protected during AI processing?" and the note beneath it reads "Ensure alignment with internal policies, regulatory requirements and customer data sovereignty obligations". **Sovereignty appears in the note, not in the question**, and the obligations it names are not specified anywhere in the document.

4.5 The shape

The document is neither of the two shapes the corpus has previously recorded.

It is **not the shape recorded of the ISM's procurement guidelines**, where the risk is named in narrative and the controls require nothing about it: here a requirement does attach, twice — to understand and ensure alignment at line 516, and to ensure alignment at line 659.

It is **not silence**: the term appears three times, a section is named for it, and a jurisdiction question is asked.

It is a third shape. **The risk is named, an activity is required, and the requirement's object is undefined.** The reader is required to align with requirements and obligations the instrument does not specify. *AI Secure by Demand* is a second instance of the same shape: an instruction to purchase only products of a kind defined elsewhere. This is reported as it stands and is not forced to either pole.

4.6 Appendix B

The document's second appendix, introduced at line 707, provides "structured questions to help organisations evaluate the effectiveness, reliability" of AI capabilities offered by cyber security vendors. Its sections are *Security outcomes and effectiveness*, *Scope, assumptions and constraints*, and *Handling uncertainty and error*.

It was read and searched for *jurisdic*, *sovereign*, *foreign*, *offshore*, *overseas*, *cross border*, *cross-border*, *compel*, *lawful*, *supply*, *supplier*, *third party*, *third-party*, *hosted*, *hosting*, *origin* and *provenance*. **Every one of those terms occurs zero times in it.** The nearest passages are a technical-control question at line 727, "What technical controls define where the AI cannot operate or influence outcomes?", and a note at line 760 that clear boundaries on data use "reduce long-term privacy, confidentiality and legal risk".

4.7 What the document does not ask

Across all twenty-four pages, *compel* and *lawful* occur zero times. The document asks where customer data is processed and stored, who supplies a capability, what its provenance and lineage are, what it depends on, and how portable it is. **It does not ask who can lawfully require production of the model, its weights, its logs, or the data it processes.**

The limits, at full width. The substring counts above are over named terms in a named file and are not concept tests. The concept test reported at section 5.4 graded this document REVIEW at phrase set 1.0.1 on two triggers, both determined and recorded there, and suppressed three bare foreign-dimension mentions as noise under its proximity rule — which is the treatment the phrase set specifies for a foreign-dimension term with no trigger beside it. Phrase set 1.0.1 did not surface the jurisdiction question at line 663, and that gap is recorded under Limitations rather than left to be found.

5 Across the population

5.1 The population and its basis

Four instruments, eight captures. The basis is the subject, the instrument the subject incorporates by reference, and the two Australian Signals Directorate publications on artificial intelligence already pinned in this corpus. Each instrument's audience is named beside it, because the audiences differ and what follows

depends on that.

Instrument	Captures	Pages	Audience, as stated	Basis for inclusion
<i>Agentic AI Harnesses — The layer above the model</i> , ASD, September 2026	AAH-1, AAH-2	12	“executive decision-makers, chief information security officers and IT leaders”	the subject
<i>Careful adoption of agentic AI services</i> , ASD’s ACSC with CISA, NSA, Canadian Centre for Cyber Security, NCSC-NZ, NCSC-UK	AAH-3, AAH-4	29	no executive audience stated; practice sections addressed to “agentic AI developers”, and to “vendors and operators”	incorporated by reference by the subject
<i>Frontier AI cyber threat considerations for boards of directors</i> , ASD with AICD, August 2026	AT-ASD-1, AT-ASD-2	7	boards of directors	pinned ASD publication on AI
<i>Opportunities for AI in cyber defence</i> , ASD with CCCS, NCSC-NZ, NCSC-UK	A-GUID-1, A-GUID-2	24	“cyber defenders, such as Chief Information Security Officers and senior security leaders”	pinned ASD publication on AI

The bound on the population is stated here and not elsewhere. The Directorate’s artificial intelligence index page was retrieved and is held at AAH-5. **Its publication list is in the markup as served and enumerates fifteen publications** — thirteen of them beneath the artificial intelligence path, each card carrying a title, a path and a description. **The population of four could have been enumerated from that index and was not.** The search that established this bound returned zero paths beneath the artificial intelligence path, and **the zero was a property of the pattern and not of the publisher**: the pattern required a quotation mark after href=, the page emits unquoted attributes, and so it matched nothing on a page listing fifteen publications. The population of four was fixed, and this assessment drafted and built, before the index was enumerated correctly on 13 September 2026; it remains four for that reason and for no other.

This assessment does not claim the population is complete, and the reason is now a choice and not an inability. The index enumerates; the eleven publications on it that this assessment does not read are named below; and it does not read them. Nothing prevented their reading except that the population had already been fixed and the document written.

What the index lists and this assessment does not read. Eleven of the fifteen publications on the Directorate's artificial intelligence index are outside this population. They are named here from the pinned index at AAH-5 and from nothing else: the index card carries a title, a path and a description, and **states no date, no length and no publisher.** None of the eleven was retrieved, and nothing below is a reading of any of them.

Thirteen of the fifteen sit beneath one directory:

/business-government/secure-design/artificial-intelligence/

They are named below by the slug that follows it. Slugs are given as plain text rather than as code, because several exceed the width of this page and plain text breaks at its hyphens where a code span does not.

- *An introduction to artificial intelligence* — an-introduction-to-artificial-intelligence
- *AI data security* — ai-data-security
- *Artificial intelligence and machine learning: Supply chain risks and mitigations* — artificial-intelligence-and-machine-learning-supply-chain-risks-and-mitigations. **Its title names supply chain, which is the subject of this assessment's central finding. It was not read.**
- *Artificial intelligence for small business* — artificial-intelligence-for-small-business
- *Content credentials: Strengthening multimedia integrity in the generative AI era* — content-credentials-strengthening-multimedia-integrity-in-the-generative-ai-era
- *Convolutional layers: An artificial intelligence primer* — convolutional-layers-an-artificial-intelligence-primer
- *Defending against AI-enabled cyber attacks* — defending-against-ai-enabled-cyber-attacks
- *Deploying AI systems securely* — deploying-ai-systems-securely
- *Engaging with artificial intelligence* — engaging-with-artificial-intelligence
- *Guidelines for secure AI system development* — guidelines-for-secure-ai-system-development. **This is the instrument *Agentic AI Harnesses* incorporates by reference. It is on the index, and it was not read.**

The two outside that directory are *Principles for the secure integration of Artificial Intelligence in Operational Technology*, at secure-design/operational-technology-environments/principles-for-the-secure-integration-of-artificial-intelligence-in-operational-technology, which this assessment does not read; and *Frontier AI cyber threat considerations for boards of directors*, which is in this population at AT-ASD-1 and AT-ASD-2 and is listed under protecting-business-leaders/cyber-security-for-business-leaders/.

5.2 What each instrument asks, and in which register

Instrument	Supply or provenance	Ownership	Location	Compellability	Register
<i>Frontier AI boards</i>	third and fourth-party supplier posture, in the threshold questions	yes , in its threshold questions and not in its action registers	no	no	threshold question; absent from all four priority registers and the practical-actions register
<i>Opportunities</i>	provenance and lineage, SBOM and AIBOM, third-party dependencies, portability	no	yes	no	vendor appendices; a requirement in <i>Identify</i> ; <i>AI Secure by Demand</i>
<i>Agentic AI Harnesses</i>	provenance assessment	no	no	no	<i>Good practice</i> , not the seven executive questions
<i>Careful adoption</i>	trusted-source origination, trusted registry, SBOM when procuring	no	no	no	<i>Best practices</i> , and the document has no question set

All four instruments reach supply. Three impose a requirement about it. One asks about ownership, in its questions and not in its actions. One asks about location, in vendor appendices. None reaches compellability.

5.3 Where the question sits, and where it does not

Three of the four instruments place their supply question in a different register from the one that tells the reader what to do.

Agentic AI Harnesses requires the model's provenance be assessed in its *Good practice* section, and asks nothing about supply, ownership or provenance in the seven questions it puts to executives.

Careful adoption of agentic AI services requires that third-party components originate from trusted sources, that a trusted registry be maintained, and that SBOM minimum elements be referenced when procuring —

all in its *Best practices* — and it has no question set at all.

Frontier AI boards asks boards about their vendors’ foreign ownership, control or influence in its threshold questions, and the terms **foreign**, **ownership** and **influence** occur zero times across the four priority registers and the practical-actions register that follow line 123, including in the register instructing boards to deploy AI models into software development, security assessment and security monitoring.

Opportunities for AI in cyber defence is the exception, and the reason is structural rather than editorial: its supply and jurisdiction questions sit in its vendor appendices, and those appendices **are** the instrument’s instruction to the reader. The document directs its audience to them from the body, at line 76: “For a list of questions to support vendor assessment, refer to Appendices A and B.” There is no separation between the register that raises the question and the register that acts on it, because the questions are the action.

In three of the four instruments, a supply question appears wherever the reader is being told to think and disappears wherever the reader is being told to act.

5.4 The negative

compel and lawful occur zero times in any of the four instruments.

The concept test is the corpus’s compellability detector at **phrase set 1.0.1, unmodified**. Every verdict reported in this assessment was produced by that version. The detector tests whether a document requires an assessment of foreign compellability, rather than whether it mentions the subject, and it requires a compulsion, instrument or reach trigger to co-occur with a foreign-dimension term in the same sentence or numbered clause.

Agentic AI Harnesses and *Frontier AI cyber threat considerations for boards of directors* each graded NO_MATCH with no concept group producing any hit. **Both have since been read end to end for synonyms the phrase set does not hold, and both are signed silent** — artefact 38f5f3b0e3b77c48 for the first, binding graded text e978c859e2d5627a, and artefact 9821a6bda1b5b1e6 for the second, binding graded text 5b32861df0597d89, each recording phrase set 1.0.1. The signoff artefact records only the graded text’s hash, so in both cases the link from the artefact to the pinned captures runs through a separate register: the retrieval log for the first, and the corpus manifest for the second.

Careful adoption of agentic AI services and *Opportunities for AI in cyber defence* each graded REVIEW, on two triggers apiece. All four triggers are the single phrase **direct**, and all four are determined: “goal-directed behaviours” and “measurable goals identified from user directions” in the first; “execute goal directed actions” and “how cyber security and cyber risk are directed, understood and managed across the organisation” in the second. Each is the governance, instruction or agentic sense. In no case did a foreign-dimension term sit within the proximity window. The determinations, their grounds and the mechanism that could not record them are held in this corpus’s open-items register at item 161.

The limits governing every figure in this section — the phrase set’s forty-one phrases and its one demonstrated gap, the fact that no verdict was produced by a modified set, the substring counts at sections 1, 3 and 4, and the bound on the population at 5.1 — are recorded under *Limitations* and are not restated here.

5.5 What this section does not claim

It does not claim that the Directorate declines to ask the compellability question anywhere. The population is the four instruments at 5.1 with the index bound stated. The Directorate publishes the Information Security Manual, technical guidance and advisories, and this corpus has separately established, in AOS-ISM-2026-001 at claims C1, C8, N2, N3 and N4, that the Manual names foreign lawful access in its narrative and requires nothing about it in the forty procurement controls or across the Manual's 1,143 control descriptions — a finding that carries its own stated limit, that the test was over control descriptions only, not narrative and not either manual end to end, and that a control could require it in words the seven terms tested do not match. The corpus's own clearest statement of the reach question is in AOS-TSC-2026-001 at claims R5 and R6, on a different subject, and it is cited here as prior work and not as support for any claim about the Directorate.

It does not claim that any omission was deliberate. Nothing on the record establishes what was considered, and this assessment does not infer it.

It does not claim that the Directorate should have asked. What an instrument ought to require is a matter for its publisher.

It does not claim that the four instruments form a pattern. A pattern claim requires a common test applied to each instrument, stated before its result is known, and that is the successor to this assessment and not part of it.

Limitations

Each limit below is a boundary of what was established. None is a defect in the instruments assessed, and none is offered as an apology. Where a limit already appears in the body it is named here with its section, so a reader meets it in both places rather than having to find it.

The population could have been enumerated from the publisher's own index and was not. The Directorate's artificial intelligence index page is held at AAH-5, its publication list is in the markup as served, and it enumerates fifteen publications. All four instruments of this population are on that list, as is *Guidelines for secure AI system development*, the instrument the subject incorporates by reference. The population of four was fixed, and this assessment drafted and built, before the index was enumerated correctly on 13 September 2026; it remains four for that reason and for no other, and the eleven publications on the list that are not read are named at section 5.1. **This assessment does not claim the population is complete, and the reason is a choice and not an inability.** Stated at section 5.1.

The bound on the population was established twice by a search that could not have found what it looked for. The search matched a quoted href pattern against a page that emits unquoted attributes; it returned zero on a page listing fifteen publications. It was run on 12 September when the bound was written, and again on 13 September when the bound was re-tested during a reading of this case's register — the same pattern, the same zero, **and the second run was reported as confirming the first.** It was found only when the same pinned bytes were searched a third time with the quotation mark requirement removed. **A re-test that repeats the method under test is not a re-test.** Nothing in this practice's checks distinguishes the two,

and nothing would have caught either run: the pinned bytes never changed, and both searches were correct programs returning correct answers to the wrong question. Recorded at item 168 of this practice's open-items register.

AAH-5 carries no logged retrieval time. The retrieval log holds four rows and the register holds five captures this practice retrieved. The record does not establish when the index page was retrieved, only that its bytes are pinned and that the enumeration over them was run wrongly twice and correctly once.

The concept test holds forty-one phrases and one demonstrated gap. The compellability detector at phrase set 1.0.1 groups its phrases by concept — compulsion act, lawful access instrument, named foreign statute, foreign dimension, reach concept — and requires a trigger to co-occur with a foreign-dimension term in the same sentence or numbered clause. It did not surface *Opportunities for AI in cyber defence*'s question "Which jurisdictions process or store customer data, including through third-party AI services?", because the group holds "foreign jurisdiction" and "overseas jurisdiction" and the document wrote the term unqualified and plural. That question is reported at section 4.1 on its own terms, having been found by reading. The test establishes that these concepts, in these phrasings and under that rule, do not occur; it does not establish that no passage carries a concept in words the set does not hold. Stated at sections 2.6, 3.6, 4.7 and 5.4.

A phrase for a later version is proposed and has not been applied. The gap above admits a narrow repair, a phrase bound to the interrogative form rather than to the bare word, which would be noisy in ordinary regulatory prose. It was not applied, and **every verdict in this assessment was produced by phrase set 1.0.1 unmodified.** Applying a changed instrument after seeing its result, and then reporting the changed instrument's output, is the practice this assessment would object to in another party.

The substring counts are not concept tests. Section 1's negative is eight stems over the full extracted text of one capture, performed twice against independent extractions. Section 3's counts are thirteen terms across a stated boundary line, with one term reported separately because it occurs on both sides in a different sense. Section 4's Appendix B negative is seventeen terms. Each was run over a named file whose hash is recorded in the references, and each establishes only that those terms do not occur.

Two instruments are signed silent, and what that asserts is narrow. *Agentic AI Harnesses* and *Frontier AI cyber threat considerations for boards of directors* each graded NO_MATCH with no concept group producing any hit, and each was then read end to end for synonyms the phrase set does not hold. The signoff artefact records the graded text's hash, the phrase set version, the signer and the statement; **it does not record which captures the graded text was extracted from.** For the first, that linkage lives in this assessment's retrieval log; for the second, in the corpus manifest of an earlier case, those captures having been pinned in August 2026. A reader following the artefact alone reaches the text and not the source. Stated at section 5.4.

Four triggers were determined by reading, and the instrument that produced them could not record the determination. *Careful adoption of agentic AI services* and *Opportunities for AI in cyber defence* each graded REVIEW on two triggers, all four being the single phrase *direct*, and in no case did a foreign-dimension term sit within the proximity window. Each was determined to be the governance, instruction or agentic sense: "goal-directed behaviours" and "measurable goals identified from user directions" in the first;

“execute goal directed actions” and “how cyber security and cyber risk are directed, understood and managed across the organisation” in the second. The detector converts NO_MATCH to silence by signed artefact and provides no equivalent record for a resolved review item, so **the four determinations and their grounds are held in this corpus’s open-items register at item i61 and nowhere else.** A reader wishing to test them reads that item.

This case’s claim register was written after its draft, contrary to this practice’s convention that a claim enters the register before it enters prose; the departure, what it costs and what stands against it are recorded at item i62 of this practice’s open-items register.

The methodological limit that governs every finding. This assessment establishes what four instruments ask and require, as at the retrieval dates in the references. It does not establish what the Directorate asks anywhere else, and the boundaries it draws on purpose — as against the limits recorded here — are set out under *What this assessment is not.*

What this assessment is not

The boundaries below were drawn on purpose. They are distinct from the *Limitations*, which record where the methods used reach their edge. A limit is what the work could not establish; a boundary is what the work did not set out to establish.

It does not assess the Information Security Manual. The Manual is a separate instrument with a separate audience, and this corpus has addressed it separately.

It does not assess any provider, vendor or supplier. No commercial party is the subject of any finding. The subjects are four publications.

It does not claim the Directorate declines to ask the compellability question anywhere. The population is the four instruments named above with the index bound stated under *Limitations*. The Directorate publishes technical guidance, advisories and the Manual, and what those require is outside this population.

It does not argue that any omission was deliberate. Nothing on the record establishes what was considered by whom, and this assessment does not infer it from what an instrument does not say.

It does not argue that the Directorate should have asked. What an instrument ought to require is a matter for its publisher. This assessment establishes what four instruments do require, and what they do not.

It does not claim the four instruments form a pattern. A pattern claim requires a common test applied to each instrument and stated before its result is known. That is the successor named under *Related published assessments*, and it is not part of this assessment.

Related published assessments

AOS-ISM-2026-001, on the Information Security Manual’s procurement guidelines, establishes at claims C₁, C₈, N₂, N₃ and N₄ that the Manual names foreign lawful access in its narrative and requires nothing about

it in its controls. C1 quotes the narrative as printed, including that “foreign-owned suppliers operating in Australia may be subject to a foreign government’s lawful access to data belonging to their customers”. C8 establishes that from “Managed services and cloud services” to the end of that chapter, no passage names foreign law, jurisdiction, offshore location, or lawful or covert collection. N2 carries the narrative forward to the September 2026 release. N3 establishes, by reading all forty controls in the chapter end to end, that none requires an organisation to establish whether a supplier is subject to a foreign government’s lawful access. N4 widens that negative across the whole Manual’s control descriptions, 1,101 rows in June and 1,143 in September, and states its own limit in the same breath: **“descriptions only, not narrative, not either manual end to end; a control could require it in words the seven terms do not match.”** That limit travels with the citation here.

That assessment’s argument is not restated. What it establishes is cited, and this assessment establishes the same shape in a different body of guidance: the instruments name and require a good deal about supply, and require nothing about who can lawfully require production. **The addition this assessment makes is that the shape appears in guidance which instructs the reader to deploy AI models into software development, security assessment and security monitoring** — that is, into the security function itself — while the supply question is absent from every register in which that instruction is given.

AOS-TSC-2026-001 is the corpus’s clearest statement of the question this population does not put. It holds at R5 *who can lawfully compel the entity that will hold the data*, and at R6 *whether the cloud is, or is not, exposed to any foreign law*. Those claims concern a Commonwealth cloud arrangement. They are cited here as prior work on a different subject, to show that the question is askable and has been asked in this corpus, and **they establish nothing about the four instruments assessed here**.

The successor

A framework for testing what a clean assurance result discloses about the population it was established over — seven dimensions: the object of the verdict, who fixes the boundary, where the boundary is recorded, who may read it, whether the published output carries the population, what consequence follows from the verdict, and how long the verdict speaks for.

That framework will be applied to Directorate instruments bearing on artificial intelligence procurement and governance, with **the Manual’s procurement chapter included as a deliberate contrast case on audience** — it is neither this audience nor this subject, and it is in the comparison for that reason.

The framework is stated here, before its result is known. This assessment does not claim its finding, does not preview its outcome, and does not depend on it.

References

Every artefact is pinned by SHA-256 in this case’s register or in the register of the case named beside it, and re-verified against its recorded hash at this case’s pin. Where this case retrieved an artefact it was fetched twice under the same agent within the minute and hashed on both fetches; one row per request is in

RETRIEVAL-LOG-2026-09-12.tsv.

- AAH-1, AAH-2 — *Agentic AI Harnesses — The layer above the model*, Australian Signals Directorate, first published September 2026 on the document's own face; landing page as served, whose *First published* field reads 11 Sep 2026, and the publisher's PDF, 12 pages. This case, 12 September 2026. Both stable across both fetches.
- AAH-3, AAH-4 — *Careful adoption of agentic AI services*, the Directorate's Australian Cyber Security Centre with the United States Cybersecurity and Infrastructure Security Agency and National Security Agency, the Canadian Centre for Cyber Security, NCSC-NZ and NCSC-UK; landing page as served, and the publisher's PDF, 29 pages, served from the publisher's May 2026 directory. This case, 12 September 2026. Both stable across both fetches.
- AAH-5 — the Directorate's artificial intelligence index page as served. Its publication list is in the markup as served and carries fifteen cards, each with a title, a path and a description. **It is evidence against the index bound this assessment first recorded, not for it:** the bound rested on a search returning zero paths beneath the artificial intelligence path, and the same pinned bytes carry thirteen. The bound as it now stands is at section 5.1 and the search that failed is recorded under *Limitations*. This case, 12 September 2026.
- AAH-T1 to AAH-T4 — the four graded texts, each extracted from the pin named on its row, and each the text over which every substring count and every detector verdict in this assessment was computed. Each detector report and each signoff artefact records the hash of the text it graded. This case, 12 September 2026.
- AT-ASD-1, AT-ASD-2 — *Frontier AI cyber threat considerations for boards of directors*, Australian Signals Directorate with the Australian Institute of Company Directors, first published August 2026 on the document's own face; publication page as served, whose *First published* field reads 05 Aug 2026, and the publisher's PDF, 7 pages. *attestation-survey-scoping-2026-08*, 5 August 2026.
- A-GUID-1, A-GUID-2 — *Opportunities for AI in cyber defence*, Australian Signals Directorate, co-sealed with the Canadian Centre for Cyber Security, NCSC-NZ and NCSC-UK; the publisher's PDF, 24 pages, and the guidance page as served. *asd-ai-cyber-defence-2026-08*, 15 August 2026.

The two silence signoffs are not register rows and are cited by hash. A register row records something retrieved or extracted from something retrieved, and carries the address it came from; a signoff artefact was generated on this practice's own machine, has no address and cannot be re-fetched by anyone. Both are held at signoffs/ in this case.

agentic-ai-harnesses-SILENCE.json binds graded text AAH-T1 at phrase set 1.0.1, signed 12 September 2026. It hashes to:

38f5f3b0e3b77c48e60a93fdde2a8c4bea3e22f47e8713594b6c3bf3766c8fcf

frontier-boards-SILENCE.json binds graded text AAH-T3 at phrase set 1.0.1, signed 12 September 2026. It hashes to:

9821a6bda1b5b1e615a3cf2ab9b823e504e2ea72fa05f2324830056e655f4103

Published assessments cited for their claims

- AOS-ISM-2026-001, claims C1, C8, N2, N3 and N4 — the Information Security Manual's procurement guidelines. `ism-procurement-2026-08`.
- AOS-TSC-2026-001, claims R5 and R6. `top-secret-cloud-sovereign-2026-09`.

The compellability detector and its phrase set are `assessments/lib/compellability_detector.py` and `assessments/lib/compellability_phrases.yaml` at version 1.0.1. The four trigger determinations the detector could not record are held in this corpus's open-items register at item 161.

End of document.