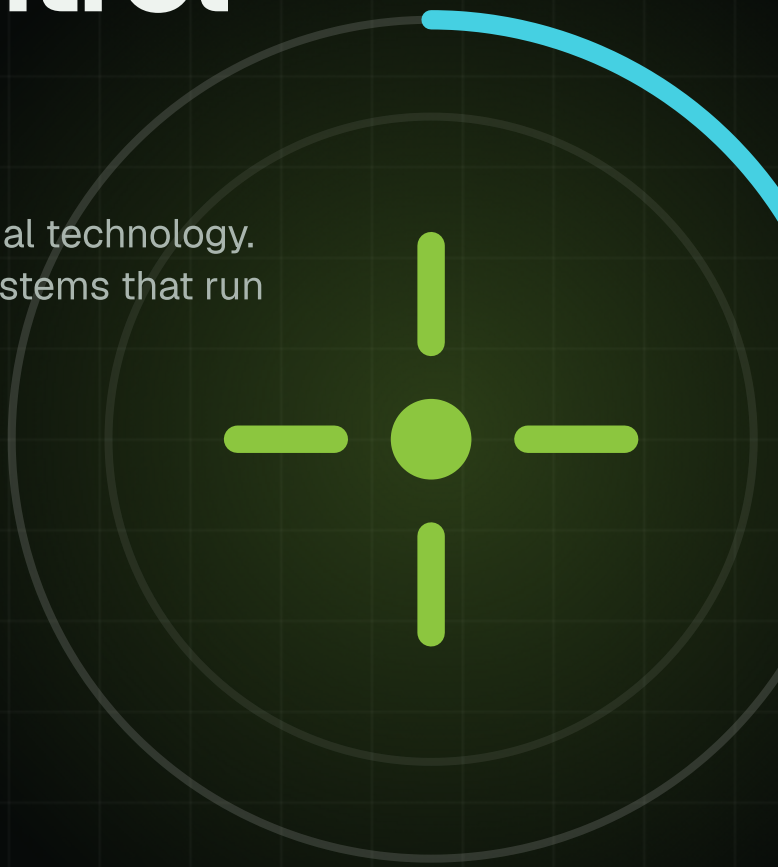
ozari

FREE BUYER'S GUIDE

AI in the control room.

A buyer's checklist for AI in operational technology. What to ask before AI touches the systems that run the grid.



36	4	8
QUESTIONS	PRINCIPLES	RED FLAGS

Based on **Principles for the secure integration of artificial intelligence in operational technology**, the joint guidance from Australia's ACSC, the US CISA and partner agencies, December 2025.

EDITION 1 · SEPTEMBER 2026
OZARI.AI

Ask better questions. Get evidence, not slides.

AI can help operators see more and decide faster. In a control room it can also add new ways to fail. This checklist turns public guidance into 36 plain questions you can put to any vendor, or to your own team.

How to use it

1. Ask each question and mark **Yes**, **Partly** or **No**.
2. Ask for the evidence in the "Ask for" column. A slide is not evidence.
3. A "No" is not always a failure. It is a risk to manage, or to accept on purpose.
4. Add up each section on page 7 and check the red flags.

☐ Yes ☐ Partly ☐ No

The four principles

1. Understand the AI

Principle 1: Understand AI · questions 1 to 7

2. Decide if AI belongs here

Principle 2: Consider AI use in the OT domain · questions 8 to 19

3. Govern and prove it

Principle 3: Establish AI governance and assurance frameworks · questions 20 to 28

4. Keep people in charge

Principle 4: Embed oversight and failsafe practices · questions 29 to 36

Who it is for. Control room managers, OT security leads, engineers and procurement teams at electricity networks, generators and other critical infrastructure.

A note on neutrality. ozari.ai builds AI for control rooms. We wrote this checklist so buyers can hold every vendor, including us, to the same standard. It paraphrases public guidance and adds some Australian context, marked "Australia". No agency has reviewed or endorsed it.

PRINCIPLE 1: UNDERSTAND AI

1 Understand the AI

Know what the AI is, how it can fail, and whether your people can work without it.

#	QUESTION, AND WHY IT MATTERS	ASK FOR	YES	PARTLY	NO
1	**What kind of AI is inside, exactly?** Statistical models, machine learning, language models and agents carry different risks.	A list of every AI component and what each one does.	☐	☐	☐
2	**What can go wrong, and how would we notice?** AI can drift, give confident wrong answers or raise false alarms.	Known failure modes, and false-alarm and missed-event rates from testing.	☐	☐	☐
3	**Does every output explain itself?** Operators need to judge advice, not just follow it.	Sample outputs with confidence and reasons, on data like ours.	☐	☐	☐
4	**Can our people run the network with the AI switched off?** Heavy reliance erodes the manual skills needed when systems fail.	A procedure for AI-off operation, and a demonstration.	☐	☐	☐
5	**Will you train our operators to question the AI?** Staff must spot invalid outputs and know what to do next.	A training plan, including checks against other measurements.	☐	☐	☐
6	**Was the AI built and maintained securely?** Secure design, procurement, deployment and operation reduce risk across its whole life.	A description of the development lifecycle and security practices.	☐	☐	☐
7	**Who is responsible for what: the vendor, the integrator and us?** Unclear roles cause confusion during an incident.	A written split of responsibilities.	☐	☐	☐

Notes

PRINCIPLE 2: CONSIDER AI USE IN THE OT DOMAIN

2 Decide if AI belongs here

Check the business case, your data, the vendor and how the AI connects to OT.

#	QUESTION, AND WHY IT MATTERS	ASK FOR	YES	PARTLY	NO
8	**What problem does it solve that simpler tools cannot?** The guidance asks owners to prefer an established capability if it meets the need.	A business case with measurable goals.	☐	☐	☐
9	**What must it achieve, and which safety and security limits must it respect?** Clear acceptance criteria stop hidden risk and scope creep.	Written success measures, and safety and security requirements.	☐	☐	☐
10	**Where is our OT data stored, processed and used for training?** Operational data and engineering files are valuable to attackers.	A data-flow diagram and a data usage policy.	☐	☐	☐
11	**Who can reach our data, from where, and under which country's laws?** Foreign laws can require a vendor to act against your interests.	A list of locations, parties and any remote or offshore access.	☐	☐	☐
12	**Will you train models on our data, or share it?** Models can keep patterns from operational data for a long time.	A contract clause that sets the rules.	☐	☐	☐
13	**Can it run on our premises, with no internet or vendor cloud?** A permanent outside connection is a new path into OT.	A demonstration on an isolated network.	☐	☐	☐
14	**Can we switch the AI features off, and who decides?** Operators need control over when the AI is active.	The switch, and what happens when it is off.	☐	☐	☐
15	**Will you give us an SBOM that covers the AI and where models are hosted?** You cannot secure what you cannot see.	A software bill of materials.	☐	☐	☐
16	**Will you tell us if you find the AI can give wrong advice or take a wrong action?** AI behaviour problems need disclosure, like vulnerabilities.	A disclosure policy that covers AI behaviour.	☐	☐	☐
17	**Does data leave OT by push or brokered paths, with no standing inbound access for the AI?** The AI must not become a way into the OT network.	An architecture showing one-way or brokered flows through the DMZ.	☐	☐	☐
18	**Can the AI change setpoints or issue commands? If so, where does a person approve?** The guidance advises limiting active AI control without a human in the loop.	The control path, and proof of the approval step, or proof there is no command path.	☐	☐	☐
19	**Will it meet our timing needs without slowing control?** AI may not meet real-time requirements.	Latency tests under realistic load.	☐	☐	☐

PRINCIPLE 3: ESTABLISH AI GOVERNANCE AND ASSURANCE FRAMEWORKS

3 Govern and prove it

Give the AI an owner, fit it into your security framework, and test it before it goes live.

#	QUESTION, AND WHY IT MATTERS	ASK FOR	YES	PARTLY	NO
20	**Who owns the AI system in our organisation?** Governance needs a named owner and support from leadership.	A named owner and a governance forum.	☐	☐	☐
21	**How does the AI fit our security framework, risk process and obligations?** AI must sit inside existing cyber and regulatory processes, not beside them.	A risk assessment. Australia: how it fits your critical infrastructure risk management program and AESCSF.	☐	☐	☐
22	**Is the AI audited regularly, including for AI-specific attacks?** Threat models should cover AI tactics, such as those in MITRE ATLAS.	The latest audit summary and threat model.	☐	☐	☐
23	**Are AI endpoints logged: flows, access and data leaving?** Logs let you detect misuse and trace decisions.	Sample logs.	☐	☐	☐
24	**How was it tested, and can we test it on our own non-production system first?** Testing should start away from production, ideally with hardware in the loop.	Test reports with method, environment and dates, and a test environment.	☐	☐	☐
25	**How do you keep production data out of test environments?** Test copies of data are a common leak.	The data handling procedure.	☐	☐	☐
26	**When a model changes, who approves it, and can we roll back?** Silent updates change behaviour without review.	The change process and version history.	☐	☐	☐
27	**Which standards does it align to, and what is certified rather than self-declared?** Claims are easy. Certificates and evidence are not.	Certificates, or a clear statement of alignment with evidence.	☐	☐	☐
28	**At what point do we fall back to non-AI operation?** Thresholds tell operators when to stop relying on the AI.	Documented thresholds and the fallback procedure.	☐	☐	☐

Notes

PRINCIPLE 4: EMBED OVERSIGHT AND FAILSAFE PRACTICES

4 Keep people in charge

Watch the AI in service, keep a person in the decision, and plan for its failure.

#	QUESTION, AND WHY IT MATTERS	ASK FOR	YES	PARTLY	NO
29	**Do we have an inventory of every AI component, and of everything that relies on it?** Oversight starts with knowing what is there.	The inventory.	☐	☐	☐
30	**Are all AI inputs and outputs logged, with the AI's identity separate from users and machines?** A separate identity makes each AI decision traceable.	A sample audit trail. Ask whether it is tamper-evident.	☐	☐	☐
31	**Where exactly do people approve AI-proposed actions?** A person in the decision improves safety, reliability and trust.	The approval workflow, demonstrated.	☐	☐	☐
32	**How do we track performance over time: drift, false alarms and missed events?** Accuracy fades as the network changes.	Performance indicators and a review schedule.	☐	☐	☐
33	**What are the safe operating bounds, and what alerts us when the AI goes outside them?** Bounds catch drift and manipulation early.	The bounds and the alerts.	☐	☐	☐
34	**Has the AI been red-teamed?** Offensive testing finds weaknesses before attackers do.	The latest red-team or penetration-test summary.	☐	☐	☐
35	**How does it fail? Can it be bypassed without disrupting operations?** New AI failure states must fail gracefully.	A failure-mode table and a demonstration.	☐	☐	☐
36	**Are AI failure states in our safety procedures and incident response plan?** Incidents will happen. Plans must include the AI.	Updated procedures and a tested plan.	☐	☐	☐

Notes

Red flags

Stop and ask again if you hear any of these.

- The AI can issue commands, and nobody can show you where a person approves.
- It needs a permanent internet or vendor cloud connection to work.
- The vendor will not say where your data goes, or whether models train on it.
- There is no way to switch the AI off, and no tested plan for running without it.
- Advice arrives without a confidence level or reasons.
- Model updates arrive silently, with no approval and no way back.
- Test results are claims, with no method, date or environment.
- "Certified" turns out to mean self-declared.

Your score

Yes = 2, Partly = 1, No = 0.

SECTION	QUESTIONS	YOUR SCORE	MAXIMUM
1. Understand the AI	7		14
2. Decide if AI belongs here	12		24
3. Govern and prove it	9		18
4. Keep people in charge	8		16
Total	36		**72**

Our advice: treat any "No" in section 4 as a stop sign until it is fixed, or until your organisation accepts the risk in writing.

Source

Principles for the secure integration of artificial intelligence in operational technology. Published 3 December 2025 by the US Cybersecurity and Infrastructure Security Agency (CISA) and the Australian Signals Directorate's Australian Cyber Security Centre (ASD's ACSC), with the NSA Artificial Intelligence Security Center, the FBI, and the national cyber security centres of Canada, Germany, the Netherlands, New Zealand and the United Kingdom.

https://www.cyber.gov.au/business-government/secure-design/operational-technology-environments/principles-for-the-secure-integration-of-artificial-intelligence-in-operational-technology

What this is, and is not

This checklist is our plain-English reading of that guidance, with some additions for Australian operators marked "Australia". It is not the guidance itself, and it is not legal or regulatory advice. It is not reviewed or endorsed by ASD's ACSC, CISA or any other agency. Read the source for the full detail.