AIR APAC

Center for AI Readiness

AIR APAC Standard · AIR APAC Pte. Ltd. (UEN 202600913G), Singapore

AIRS-J/1

VERSION 2.1 · IN FORCE

Judgment in AI-Enabled Work

This Standard defines the competency model, rubric, evidence requirements and decision rules used by AIR APAC to certify individual judgment in AI-enabled work.



Dated 2026-10-15. Released 2026-10-15. airapac.org/standard/v2-1. © 2026 AIR APAC.

Published in full, so any party can inspect the criteria

Status of this edition. Version 2.0 moves AIRS-J/1 from a published specification to an operating certification. AIR APAC is the judgment authority: it defines what good judgment in AI-enabled work looks like, assesses people against it through its official instrument, and stands behind every result: it publishes the rules, gives the reasons for every level, and corrects a result where an appeal succeeds (Section 18.5). The evidence programme in Annex A runs alongside certification and strengthens it with every assessment; it no longer holds certification back. Compatible refinements are issued as v2.x, material changes to the competency model, anchors or decision rules as v3.0.

Contents

	At a glance	3
	The Standard: document control	5
1	Purpose	6
2	Scope	7
3	Terms	9
4	Governance and independence	10
5	Competency model	11
6	Stakes profile and comparability control	12
7	Proficiency scale and anchors	13
8	Assessment methods and evidence	19
9	Assessment integrity	20
10	Confidence	21
11	Reliability, standard setting and automated scoring	21
12	Certification decision rules	24
13	Tracks	27
14	Register, credential portability and data	29
15	Fairness monitoring	29
16	Relationship to other AIR APAC instruments	30
17	Governance roster	31
18	Recertification, suspension and appeal	31
19	Version history and change control	32
20	Training aligned to the Standard	34
A	Evidence programme	37
B	Crosswalk and retirement	39
C	Assessor observation record	41
D	Critical-review patterns	42
E	Informative global benchmarks	43
F	The official instrument	44

9

Dimensions of judgment, grouped under four STEP domains (Section 5)

0 to 4

The proficiency scale every observation is scored on (Section 7)

6

Certification thresholds, every one of which must be met (Section 12.2)

© 2026 AIR APAC. All rights reserved.

INFORMATIVE · SECTION 5 GOVERNS

Nine dimensions, grouped under four STEP domains.

Structure · Framing the problem before reaching for a tool.

- S1 Problem Framing
- S2 Scope and Constraint Definition

Test · Interrogating evidence, including machine-generated evidence.

- T1 Evidence Verification
- T2 Assumption and Alternative Testing
- T3 Risk and Failure Anticipation

Engage · Authority, people, and being seen to act.

- E1 Decision Rights and Escalation
- E2 Stakeholder Alignment and Disclosure

Perform · Deciding proportionately, and owning the decision.

- P1 Proportionate Commitment Under Uncertainty
- P2 Consequence Ownership and Correction

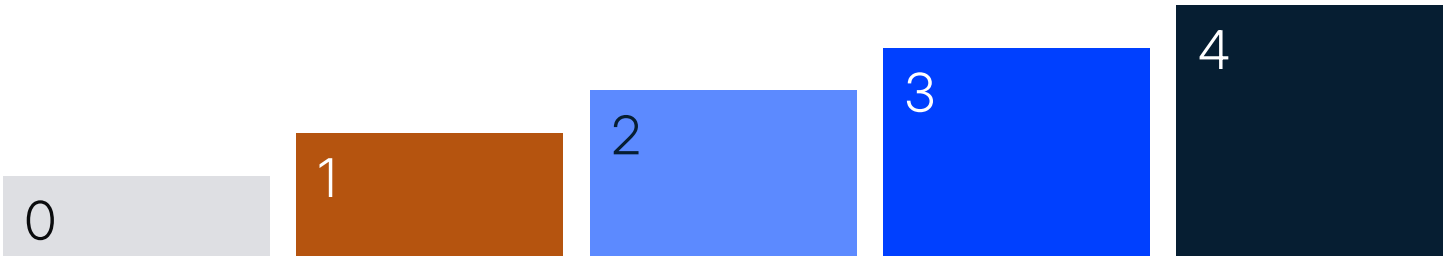


Nine dimensions is a reporting hypothesis, not an established factor structure. The model is designed so it can shrink to four domains without breaking the certification decision rules. It will shrink if the evidence says so.

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

INFORMATIVE · SECTIONS 7 AND 12 GOVERN

Every observation is scored 0 to 4.



Not demonstrated

The behaviour was absent, or its opposite was displayed.

Emerging

The behaviour appeared partially, late, or only after prompting.

Competent

The behaviour appeared unprompted and was adequate to the situation. Threshold level.

Proficient

As Level 2, calibrated to consequence, and made inspectable: recorded or stated so another person could audit the reasoning.

Exemplary

As Level 3, and the judgment is made transferable: the candidate leaves a reusable rule, condition, indicator or check that would govern the next decision of this class.

All six must be met.

01 Dimension floor

No dimension below 2.0

02 Domain floor

Each of S, T, E, P at or above 2.0 (mean of its dimensions)

03 Composite

Overall mean at or above 2.3 (Section 11.1)

04 Evidence

Section 8.3 satisfied for all nine dimensions

05 Stakes coverage

Section 6.4 satisfied

06 Confidence

Moderate or High

2.3 – 2.9

Certified. All floors met

3.0 – 3.4

Certified with Merit. No dimension below 2.5

3.5 and above

Certified with Distinction. No dimension below 3.0

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

AIR APAC Standard for Judgment in AI-Enabled Work

Designation	AIRS-J/1
Version	2.1
Status	Approved by the owner of AIR APAC on 2026-10-15; the review board was told under Section 4.2. Released together with the permanent URL https://airapac.org/standard/v2-1. It supersedes v2.0, which keeps its permanent URL https://airapac.org/standard/v2-0. Certification is open.
Issuing body	AIR APAC Pte. Ltd. (UEN 202600913G), Singapore
Effective date	2026-10-15 (v2.0: 2026-10-05; v1.1: 2026-10-01; v1.0: 2026-08-19)
Review cycle	Annual, with interim errata permitted
Public rubric	This document is published in full, with the official instrument's specification incorporated by reference in Annex F. No scoring criterion applied in certification exists outside these two documents.

Status of this edition. Version 2.0 moves AIRS-J/1 from a published specification to an operating certification. AIR APAC is the judgment authority: it defines what good judgment in AI-enabled work looks like, assesses people against it through its official instrument, and stands behind every result: it publishes the rules, gives the reasons for every level, and corrects a result where an appeal succeeds (Section 18.5). The evidence programme in Annex A runs alongside certification and strengthens it with every assessment; it no longer holds certification back. Compatible refinements are issued as v2.x, material changes to the competency model, anchors or decision rules as v3.0.

Rendered from **`airs-j-1-v2-1.md`** (SHA-256 5c115a1667d2). Canonical source: AIR_APAC/standards/AIRS-J-1.md. This PDF edition was produced from the web edition at airapac.org/standard/v2-1 on 23 October 2026; check that address for errata issued since.

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

1 Purpose

This Standard defines the competency model, rubric, evidence requirements and decision rules used by AIR APAC to certify individual judgment in AI-enabled work.

It exists so that:

1. A certified individual holds a portable credential whose meaning does not depend on their employer.
2. Two individuals assessed in different organisations, countries or delivery settings receive comparable results.
3. Any party (a candidate, an employer, a regulator, a competitor) can inspect the criteria against which a decision was made.
4. An employer can rely on an expert, independent reading of a person's judgment when it develops, assesses, promotes or hires them.

Comparability is the governing design constraint. Where customisation and comparability conflict, this Standard resolves in favour of comparability. Where comparability and evidential honesty conflict, this Standard resolves in favour of evidential honesty: a result that cannot be defended is worse than no result.

1.1 The problem this addresses

AI systems now produce the answer. They do not carry the consequence of acting on it. The scarce and consequential human contribution has shifted from producing work to deciding whether work is fit to act on, who is entitled to be told, what may not be delegated, and what happens when it turns out to be wrong.

Organisations have no common language for that contribution, and no way to compare it across two people, two teams or two markets. Training attendance, tool proficiency certificates and self-reported confidence do not measure it. This Standard measures it, and AIR APAC certifies it.

Managers are often reluctant to judge their people's judgment, because the consequences of getting it wrong fall on them. This Standard gives them a published, evidence-based reading to rely on.

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

2 Scope

In scope. Assessment and certification of individual human judgment when working with, alongside, or in supervision of AI systems in a professional or pre-professional setting.

Out of scope for this version.

- Organisational readiness rating (see Section 16.1)
- Vendor and product certification (see Section 16.2)
- Technical proficiency with any specific AI tool, model or vendor
- Domain knowledge in any professional field

In scope from v2.0: use of results for development, employee assessment, promotion and hiring (Section 2.1), and team-level reporting to the organisation that commissions an assessment (Section 12.7).

Three uses, one rubric (from v2.1). The same dimensions, anchors, thresholds and decision rules serve three uses. They differ only in what evidence is gathered, who receives it and what it may decide.

Use	Who it serves	Evidence	What it may decide	Governing sections
Development	Adult workers in training programmes	Practice and session readings	Nothing about a person's employment or certification	10.1, 12.8, 20
Readiness	Final-year students and recent graduates	Campus track Assessment	Campus certification, released to employers by the holder, or used by an employer that commissions a Campus-track Assessment of its own graduate applicants (Section 13.4)	13, 13.4, 13.5
Assessment for employment	Employers assessing, promoting or hiring	Assessment at Moderate or High confidence	Certification, and AIR APAC's expert reading for the employer's decision	12.6, 12.8 to 12.10, 15.7

This Standard assesses **how a person decides**, not what they know.

2.1 What a result and a credential claim

These claims are normative. Each rests on a named basis in this Standard, and AIR APAC, its delivery partners and its licensees state them in those terms.

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

A result or credential is	Basis
Validated	It meets the validation programme in Section 11.7, checks V1 to V5, as published in the current validation report.
Psychometric	Levels are derived by published rules from observed evidence, with reliability and stability shown under Section 11.7, checks V2 and V3.
Predictive of job performance	Higher levels go with better rated performance at work, in a criterion study on real people in a named role family, published in the validation report (Section 11.7, check V8). The claim names the role family and the study, and says whether the study was concurrent or predictive. Prediction of unseen assessment situations (V6) supports it and is not its basis alone. Until V8 is published for a role family, the claim for that role family is the "For hiring, promotion and employee assessment" row below.
Bias-free	Matched candidates who differ only in gender, age, nationality, first language, market or job level reach the same levels (Section 11.7, check V7), and outcomes are monitored under Section 15, including adverse impact on real applicants where results are used for hiring (Section 15.7).
For hiring, promotion and employee assessment	AIR APAC's expert reading of the person's judgment, which the employer may rely on, given only from an employment-grade result (Section 12.8). Every level traces to the person's own acts and a published rule (Sections 8.8, 11.4). The employment decision is the employer's (Section 12.6).
Judgment-ready for AI-enabled work (Campus)	The holder met the certification rules on the Campus track, on workplace scenarios at entry level (Section 13).
Aligned to AIRS-J/1 (a training programme)	The programme meets Section 20.1 and publishes its outcome map. It is a claim about the programme, not about any participant.

What it is not, because it measures something else:

The credential is not	Because
A measure of intelligence or personality	The construct is observable decision behaviour, not a trait.
A measure of AI technical skill	Prompt craft, model selection and tool fluency are not assessed.
A measure of domain expertise	Scenarios supply the domain content the candidate needs.
A substitute for supervision, professional licensing or regulatory authorisation	Nothing in this credential authorises a holder to act beyond their existing professional or legal authority.

A holder or organisation that states a claim beyond its basis in this Standard is acting outside it.

2.2 Certified function and use of "managerial"

This Standard certifies the exercise of judgment in AI-enabled work using domain content supplied within an approved assessment. It does not certify occupational competence, domain expertise, comprehensive AI literacy, management of people, or performance in an unbounded workplace.

In this Standard, **managerial** describes management of decisions, consequences and AI outputs. It is a construct label, not a claim that the holder occupies, or is qualified to occupy, a people-management role.

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

The plain-language description on every credential and register entry must carry this scope. Track remains part of the credential meaning under Section 13.

3 Terms

Term	Definition
Dimension	One of nine discrete judgment competencies defined in Section 5. In this Standard, "dimension" always refers to an individual judgment dimension.
Domain	One of four groupings of dimensions under the STEP structure: Structure, Test, Engage, Perform.
Observation	A single scored instance of one dimension, arising from one assessed decision point, supported by identified evidence in the transcript.
Decision point	A moment in an assessment at which the candidate must commit to an action, position or artifact that carries cost.
Anchor	The behavioural description defining a proficiency level for a dimension (Section 7).
Stakes profile	The consequence families in play in a given assessment instance (Section 6).
Confidence	A statement of evidential sufficiency attached to every result (Section 10).
Access support	An approved adjustment to presentation, timing or response modality that removes a construct-irrelevant barrier without directing the candidate toward an assessed behaviour.
Approved instrument	The official instrument (Section 16.3, Annex F), or an instrument approved against Sections 8, 9 and 10 and the criteria in Annex A12. An eligible instrument type is not itself an approved instrument.
Managerial	Managing decisions, consequences and AI outputs; not necessarily managing people (Section 2.2).
Transcript	The complete retained evidence record of an assessment instance, from which every observation must be independently derivable.
Employment-grade result	An Assessment result from an approved instrument at Moderate or High confidence, or a valid credential shown by its holder (Section 12.8).
Development reading	A reading from practice, a training session or a result below Moderate confidence. It shows decisions and how the published rules read them, carries no level, band or confidence, and decides nothing about employment or certification (Sections 10.1, 20.3).
Licensed trainer	A trainer who holds a written licence from AIR APAC or AIR Academy to deliver a programme aligned to this Standard (Section 20).

Naming note. The AIR APAC AI Readiness Index uses "dimension" for its organisational public-signal components. Those are referred to in this Standard as **Index signal components** to avoid collision. The two models share no structure and are not convertible.

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

4 Governance and independence

These provisions are normative. They constrain AIR APAC, not the candidate.

4.1 **Published rubric.** The complete rubric is public and versioned. AIR APAC applies no undisclosed criterion, weighting, adjustment or override. Any criterion not in this document, or in the instrument specification it incorporates (Annex F), does not exist.

4.2 **Review board.** The review board is drawn from AIR APAC's standing councils, published at airapac.org/about (Section 17). It reviews every major version before release and advises on the evidence programme (Annex A). The owner of AIR APAC decides, and holds a veto over the board's recommendations. Each board member's dissent from a released version is recorded in the version history.

4.3 **Training and certification.** Training is how most candidates meet this Standard: AIR APAC, AIR Academy and certified trainers deliver it, and the official instrument offers practice. Certification outcomes are decided by the rules in Section 12 alone, never by what a candidate or employer bought. The owner of AIR APAC takes no part in any individual result. A candidate may present for assessment without prior training. Practice content and certifying content are kept apart (Section 9.3).

4.4 **No outcome-contingent fees.** No assessor, delivery partner, trainer or licensee is paid on a basis that varies with candidate outcomes. Partner fees are per participant or per session. No trainer, delivery partner or licensee receives a fee or referral payment that varies with whether a participant presents for, or obtains, certification.

4.5 **No customer-authored criteria.** Delivery partners and licensed trainers may not vary, extend, reweight or supplement the dimensions, anchors, thresholds or decision rules in this Standard. Contextual localisation of scenario content is permitted under Section 8.5; localisation of criteria is not.

4.6 **Assessor independence.** Where a human assessor scores or confirms an observation, no individual may act as sole assessor for a candidate they have personally trained within the preceding twelve months, nor for a candidate within their own reporting line, nor where they hold a financial interest in the outcome. Assessors declare conflicts before each assessment; declarations are retained.

4.7 **Public register.** Every valid certification is recorded in a verifiable register (Section 14).

4.8 **Findings act on the Standard.** Where reliability, fairness or validity monitoring shows a claim needs strengthening, AIR APAC acts on it in the next release: it fixes the instrument, the rule or the anchor, and the validation report records what changed.

4.9 **Declared interests of the issuing body.** The owner of AIR APAC Pte. Ltd. also owns Exeter Labs. StepSim is an AIR APAC initiative, built and operated by Exeter Labs, and is the official instrument of this Standard (Section 16.3). Exeter Labs also holds the intellectual property in, builds and operates the other tools AIR APAC uses, under contract with AIR APAC. AIR Academy, AIR APAC's learning division, delivers the StepSim facilitator programme under a written arrangement with Exeter Labs. Exeter Labs issues the facilitator certification; it is a delivery credential, not a certification under this Standard. Any further arrangement between the two is entered here before it takes effect.

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

5 Competency model

Nine dimensions, grouped under four STEP domains.

The model is derived from the STEP judgment construct already canonical in AIR APAC's simulation methodology, and is cross-walked to the scenario-authoring lifecycle in Annex B. It supersedes all earlier competency structures, including the ten-domain readiness model, which is retired (Annex B.3).

S Structure

Framing the problem before reaching for a tool.

S1 Problem Framing
Identifies the actual decision to be made now, distinguishing it from the presenting request, the symptom, or the task as delegated. Identifies what part of the responsibility cannot be delegated.

S2 Scope and Constraint Definition
Establishes what is in bounds, what is fixed, what is negotiable, what is unknown, and what information would be sufficient to proceed.

T Test

Interrogating evidence, including machine-generated evidence.

T1 Evidence Verification
Traces claims to source. Distinguishes generated, retrieved, asserted and verified content. Applies verification effort proportionate to consequence.

T2 Assumption and Alternative Testing
Surfaces the assumptions a decision rests on, names competing explanations, and takes action capable of discriminating between them, including against their own preferred position.

T3 Risk and Failure Anticipation
Identifies plausible failure modes, the pathway to each, who bears the cost, and which harms are being accepted rather than mitigated.

E Engage

Authority, people, and being seen to act.

E1 Decision Rights and Escalation
Acts within their authority, escalates what requires another party's approval, and does not delegate what cannot be delegated, to a person or to a machine.

E2 Stakeholder Alignment and Disclosure
Identifies who must be consulted before a decision holds and who must be informed after; discloses AI

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

involvement, uncertainty, limitation and self-interest to parties entitled to know, without prompting; communicates so the recipient can act.

Facets, retained for authoring and assessor clarity, not separately reported: E2a consultation and sequencing; E2b disclosure of AI use, uncertainty and interest; E2c audience-appropriate communication.

P Perform

Deciding proportionately, and owning the decision.

P1 **Proportionate Commitment Under Uncertainty**
Commits to a defensible position on incomplete information within the time available, at a cost and reversibility proportionate to the evidence and the stakes, and states the basis for it.

P2 **Consequence Ownership and Correction**
Accepts the outcome without transferring it to the AI system, the data or another party. Holds, modifies or reverses a prior position on the merits when evidence changes, and distinguishes a poor decision from an unlucky outcome.

5.1 Why nine, and the condition under which it becomes fewer

Nine dimensions is a **reporting hypothesis**, not an established factor structure. The minimum evidence requirement in Section 8.3 yields 27 observations; that is sufficient to score nine dimensions but not sufficient to demonstrate that nine dimensions separate empirically.

Monitoring rule. All nine dimensions are reported. The evidence programme (Annex A) monitors agreement (Section 11.2) and separation (Section 11.3); a dimension that does not separate from another in its domain is merged in the next major version, and this Standard says so.

The model is designed so it can shrink to four domains without breaking the certification decision rules. It will shrink if the evidence says so.

Known separation risks, stated in advance: T1/T2 (both interrogate inputs); S1/S2 (S2 may prove a facet of framing rather than a dimension); E1/E2 (both concern parties outside the candidate).

6 Stakes profile and comparability control

Judgment scores are not comparable across situations of different consequence. A Level 3 verification in a low-stakes exercise and a Level 3 verification where a person could be harmed are not the same behaviour.

6.1 Every assessment instance declares a **stakes profile**: one primary consequence family and at most one secondary, drawn from the canonical set below.

Code	Consequence family
C1	People safety and welfare
C2	Financial and fiduciary loss

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

Code	Consequence family
C3	Operational and service continuity
C4	Information and privacy harm, including wrong or misused records about identifiable people
C5	Trust, reputation and legitimacy (C5a commercial; C5b institutional and public)
C6	Environmental and community-system harm

6.2 Consequence families are **not scored**. They are recorded so that observations are compared within like stakes.

6.3 **Eligibility.** A consequence family belongs in a stakes profile only where the instance identifies an affected bearer, provides a credible pathway from the candidate's decision to the harm, makes severity or reversibility relevant, and requires the candidate to accept, reduce, transfer or monitor that harm. Mentioning a policy, a dataset or a stakeholder is not sufficient.

6.4 **Coverage requirement.** A certifying assessment must span at least **three distinct primary consequence families**, of which at least one must involve irreversible or non-financial harm (C1, C4, C5b or C6). A candidate assessed only on financial and operational scenarios has not demonstrated transferable judgment and is scored Insufficient Evidence.

6.5 **Contextual qualifiers.** Duty of care and vulnerability, public accountability, dependency and cascade, and binding legal or contractual constraint may be recorded as qualifiers on an instance. They are not scored and are not consequence families.

6.6 **Reporting.** A candidate's report states the stakes profiles under which their observations were generated. An employer reading a Certified result can see the consequence range it was earned across.

7 Proficiency scale and anchors

Every observation is scored 0 to 4.

Level	Label	General meaning
0	Not demonstrated	The behaviour was absent, or its opposite was displayed.
1	Emerging	The behaviour appeared partially, late, or only after prompting.
2	Competent	The behaviour appeared unprompted and was adequate to the situation. **Threshold level.**
3	Proficient	As Level 2, calibrated to consequence, and made inspectable: recorded or stated so another person could audit the reasoning.
4	Exemplary	As Level 3, and the judgment is made transferable: the candidate leaves a reusable rule, condition, indicator or check that would govern the next decision of this class.

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

7.1 **Prompting.** Where an instrument scaffolds or prompts a behaviour, the scaffolding level is recorded in the transcript. A prompted response caps at Level 1 for that observation. Guided practice may scaffold; certification may not silently roll a scaffolded response into a comparable total.

7.1.1 **Access support is not prompting.** Prompting under Section 7.1 means content-directed scaffolding toward the assessed behaviour. An approved access support is recorded in the transcript but imposes no level cap. It may change presentation, timing or input and output modality only to remove a construct-irrelevant barrier. It may not reveal a decision path, criterion, consequence or expected response.

7.1.2 **A response bank is not prompting.** An instrument may let candidates build their work product by choosing from a bank of authored lines, where the same optional bank is offered at every comparable decision point, includes plausible lines that are wrong or irrelevant, and does not signal which lines are expected. Choosing a line from such a bank is the candidate's own unprompted statement for Sections 7.1 and 7.5. Lines are offered only where the instrument specification (Annex F) says so.

7.2 **Level 4 is not seniority.** Level 4 requires only that the candidate articulate a rule that would carry forward. It does not require organisational authority to impose it, does not require the rule to be adopted, and is therefore attainable on every track including Campus. Any anchor that would require positional power to satisfy is defective and must be rewritten.

7.3 **No intent inference.** Anchors describe what appears in the transcript. An assessor may not score on inferred motivation, attitude, confidence or personality. If an anchor cannot be applied from the transcript alone, it is defective.

7.4 **Levels are cumulative.** Level n requires everything in Level n−1 to be satisfied. An impressive Level 4 behaviour without the Level 2 base is scored at the highest level whose requirements are fully met.

7.5 Anchor sets

S1 Problem Framing

Level	Anchor
0	States no decision, or restates the presenting request as the decision. Proceeds to act on the delegated task as given.
1	Names a decision, but it is the delegated task rather than the underlying decision; or arrives at the correct framing only after the situation forces it.
2	States, before investigating, the decision that must be made now, who owns it, and what it is not. Separates the presenting request from the underlying decision.
3	As Level 2, and states what part of the responsibility cannot be delegated, to another party or to an AI system, and records the frame in a form another person could review.
4	As Level 3, and names the condition that would make this a different decision rather than a revised one: a stated trigger for reframing.

Common Level 0 pattern: accepting "summarise this and recommend an option" as the decision, when the decision is whether the option set is the right one.

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

S2 Scope and Constraint Definition

Level	Anchor
0	Proceeds without identifying any boundary. Treats all information as available and all options as open.
1	Repeats the constraints given in the brief. Does not identify unknowns, or does not distinguish what is fixed from what is negotiable.
2	Names, unprompted, what is in scope, what is fixed, and at least one material unknown that must be resolved before deciding.
3	As Level 2, and distinguishes fixed from negotiable constraints, and states what information would be **sufficient** to proceed, not merely what is missing.
4	As Level 3, and states a stopping rule in advance: the point at which information gathering ends and the decision is made on what is held.

Level 3 distinction: listing every gap is not Level 3. Level 3 requires knowing which gaps would change the answer.

T1 Evidence Verification

Level	Anchor
0	Acts on an unverified claim where verification was available and the claim was load-bearing. Treats model or system output as established fact.
1	Expresses doubt about a claim but acts on it unchanged; or verifies only after being challenged.
2	Checks the load-bearing claim against an available source before acting, unprompted. Distinguishes generated content from retrieved content.
3	As Level 2, with verification effort proportionate to consequence: heavy where the decision is irreversible or high-consequence, light where it is neither. States in the work product which claims were verified and which were not.
4	As Level 3, and specifies a check that persists beyond this decision, so the same class of unverified claim is caught next time.

Level 0 also covers the inverse: rejecting AI output wholesale without examining it is a failure to verify, not caution.

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

T2 Assumption and Alternative Testing

Level	Anchor
0	Treats the first available explanation as the explanation. No alternative is named or considered.
1	Acknowledges that other explanations exist but does not name one, or names one without testing it.
2	Names, unprompted, at least one assumption the decision depends on and at least one competing explanation, and takes an action capable of discriminating between them.
3	As Level 2, and states what evidence would falsify their own preferred position, then acts on that evidence when it appears.
4	As Level 3, and tests an assumption they themselves introduced, not only one embedded in the brief, the data, or the AI output.

Level 3 is the discriminating level. Naming a falsifier and then ignoring it when it arrives scores 1, not 3.

T3 Risk and Failure Anticipation

Level	Anchor
0	Names no failure mode, or names only the risk of not acting.
1	Names a generic risk without identifying how it would occur or who would bear it.
2	Names, unprompted, at least one plausible failure mode, the pathway by which it would occur, and who carries the cost.
3	As Level 2, and distinguishes reversible from irreversible failure, and states which harm is being **accepted** rather than mitigated.
4	As Level 3, and names a leading indicator (an observable that would appear before the failure) and who is responsible for watching it.

Level 3 note: naming an accepted exposure is not a weakness in the answer. Failing to name it is.

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

E1 Decision Rights and Escalation

Level	Anchor
0	Acts outside their authority, or delegates a decision that cannot be delegated, to a person or to an AI system, without acknowledgement.
1	Recognises an authority boundary but does not act on it; or escalates everything, using escalation to avoid deciding.
2	Acts within their authority, unprompted, and escalates the parts requiring another party's approval, identifying who that party is.
3	As Level 2, with the split proportionate: escalates what genuinely requires it, decides what does not, and states the basis for the split.
4	As Level 3, and names the condition under which the boundary itself should be reset for future cases, rather than only complying with it here.

Both failure directions score 1 or 0. Over-escalation is not safe behaviour; it transfers the decision without transferring the reason.

E2 Stakeholder Alignment and Disclosure

Level	Anchor
0	Omits a party entitled to be consulted or informed; or conceals AI involvement, material uncertainty, a limitation, or a personal interest from a party entitled to know.
1	Identifies affected parties but engages them after the decision is fixed; or discloses only when asked.
2	Identifies, unprompted, who must be consulted before the decision holds and who must be informed after; discloses AI involvement and material uncertainty in the work product.
3	As Level 2, with communication fitted to each recipient's decision, so the recipient can act without a further round of clarification, and competing interests named rather than smoothed.
4	As Level 3, and frames the disclosure as a standing expectation for this class of work rather than a one-off courtesy.

Disclosure of AI involvement is assessed at Level 2, not as a bonus. Its absence, where a recipient would reasonably want it, is Level 0.

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

P1 Proportionate Commitment Under Uncertainty

Level	Anchor
0	Does not commit; commits after the window has closed; or commits to a position the available evidence contradicts.
1	Commits, but the response is not proportionate to the stakes (maximal caution where cost is low and harm minor, or minimal response where harm is severe or irreversible), or commits without stating a basis.
2	Commits, unprompted and within the time available, to a position whose cost and reversibility are proportionate to the evidence and the stakes, and states the basis.
3	As Level 2, and states in advance what would cause the position to change: a named condition, not a general willingness to reconsider.
4	As Level 3, and the committed position is executable by the named recipient without further instruction: owner, first action and timing specified.

Proportionality cuts both ways. Recommending the most cautious option regardless of stakes is not judgment; it is the absence of it.

P2 Consequence Ownership and Correction

Level	Anchor
0	Deflects the outcome to the AI system, the data, another party, or the brief. Or holds a position that new evidence has disproved.
1	Accepts the outcome in general terms without identifying what specifically was wrong; or revises only under pressure rather than on evidence.
2	On new evidence, holds, modifies or reverses the position on the merits and states which evidence drove the choice. Accepts the consequence of the original decision without transferring it.
3	As Level 2, and distinguishes a poor decision from an unlucky outcome: corrects reasoning that was wrong, and does not abandon reasoning that was sound but unlucky.
4	As Level 3, and identifies what in their own process produced the error, in terms that would change how the next decision of this class is made.

Holding a sound position under challenge is Level 2 behaviour. Changing position because someone senior pushed back, absent new evidence, is Level 1.

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

8 Assessment methods and evidence

8.1 **Instruments.** Observations for certification are generated through an approved instrument meeting Sections 8, 9 and 10. **StepSim is the official instrument** (Section 16.3): a scenario simulation with retained decision trace, specified in Annex F. Further instruments may be approved against the same sections. Eligible instrument types are:

- scenario simulation with retained decision trace;
- assessed live exercise with an authorised assessor;
- structured work-product review against a defined brief.

8.2 **Self-report is not evidence.** Self-assessment, survey response, declared experience, training attendance and manager endorsement generate no observations and contribute nothing to a score.

8.3 **Minimum observations.** Each dimension requires a minimum of **three independent observations** arising from distinct decision points. Two observations from the same decision point are one observation. A dimension below the minimum is scored Insufficient Evidence and cannot be certified.

8.4 **Assessment volume.** The minimum implies 27 observations. The official instrument's coverage per scenario, and the number of scenarios a certifying assessment needs, are set in its specification (Annex F). StepSim's rule is at least three held-back Assessment scenarios.

8.5 **Contextual localisation.** Scenario setting, sector, role titles, organisational norms, regulatory context and language may be localised. The decision points, the dimensions they elicit, the stakes profile and the anchors applied may not. Localised scenarios require AIR APAC approval before use in certification.

8.6 **Decision-point integrity.** Where an instrument reveals consequence, the candidate's position must be recorded **before** revelation. Post-revelation revision is an observation under P2, not a correction to P1. An instrument that permits silent overwriting of a committed position is not approvable.

8.7 **Evidence.** Transcripts retain the full decision trace: decisions, artifacts, sources consulted, stated rationale, revisions and timings. Keystroke capture, affect inference, video analysis and biometric signals are never scoring inputs: the Standard reads what a person did, which can be shown to them and explained.

8.8 **Derivability.** Every observation must be traceable to identified evidence in the transcript. An observation whose evidence cannot be pointed at does not exist.

8.9 **Accessibility and accommodations.** Every certifying instrument meets published accessibility requirements (set for StepSim in Annex F) and supports a documented accommodations process. A candidate may request an access support before assessment and receives a written decision with reasons. Approved supports must preserve the construct, decision points, stakes profile and assessment integrity; they are recorded under Section 7.1.1 and Annex C. A denial or material misapplication is appealable under Section 18.6. Request procedures and decision authority are published with this edition's release.

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

9 Assessment integrity

A simulation-based credential fails at the point where scenarios leak or identity is unverified. These provisions are load-bearing.

9.1 **Identity and sitting mode.** Candidate identity is verified at the start of each certifying session by a method recorded in the assessment log. Remote sessions require a live check. An Assessment commissioned for a hiring decision (Section 12.6, Applicants) is sat in person under supervision. Remote sittings for hiring open only when the approved instrument records the session (screen and camera) for its duration; AIR APAC announces this on its list of approved instruments (Section 16.3). A recording is integrity evidence under Section 18.4 and is never a scoring input (Section 8.7). Every employment-grade result states the sitting mode.

9.2 **AI use during assessment.** The AI systems provided within the assessment environment are **expected to be used**; declining to use them is not a virtue and is not scored as caution. Use of AI tools **outside** the provided environment during a certifying session is prohibited and constitutes misrepresentation under Section 18.4. This is stated to candidates before the session begins.

9.3 **Scenario exposure and retirement.** Each certifying scenario carries an exposure counter. A scenario is retired from certification use when exposure exceeds the published threshold, when it appears in public preparation material, or when performance distribution shifts in a manner consistent with content leakage. Retired scenarios may be released for practice. AIR APAC publishes the list of released practice scenarios, so that trainers and institutions know which content may be used in practice.

9.4 **Parallel forms.** Each track maintains at least two non-overlapping certifying forms of equivalent stakes coverage. A candidate reassessing after a Not Certified outcome receives a different form. A candidate may sit again no sooner than 30 days after a Not Certified outcome. An organisation that commissioned an assessment receives every result of the sittings it commissioned.

9.5 **Item leakage monitoring.** AIR APAC monitors for published solutions and coaching content and treats confirmed leakage as a retirement trigger, not a marketing problem.

9.6 **Version pinning.** Every transcript records the scenario version, rubric version, instrument version and scaffolding level. A result that cannot state these is void.

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

10 Confidence

Every result carries a confidence statement. Confidence describes **evidential sufficiency**, not the candidate's ability.

Confidence	Condition
High	Five or more observations per dimension, from four or more scenarios, scored by rules that meet Section 11.2. Stakes coverage under 6.4 exceeded.
Moderate	Three to four observations per dimension, from three or more scenarios, stakes coverage under 6.4 met.
Provisional	Minimum count met, but observations cluster in a single scenario, or stakes coverage under 6.4 is not met. Several scenarios in one sitting are not clustering.
Insufficient	Minimum not met for one or more dimensions. Reported as a development reading, with the observation count per dimension; no band.

Certification may be issued at **Moderate or High only**. Provisional results are reported to the candidate and are not certification. Insufficient results produce no certification and no band.

10.1 **Precedence.** Conditions are evaluated in the order Insufficient, Provisional, Moderate, High. If any Provisional condition is met, the result is Provisional regardless of observation count. Insufficient and Provisional results are development readings: they are reported to the candidate and, where the assessment was commissioned by the candidate's employer, to the organisation for development only (Section 12.6, Employees). Where the assessment was commissioned for a hiring decision, Section 12.6, Applicants applies and the employer is told only that no employment-grade result was obtained. An Insufficient result shows how many observations were obtained per dimension, not a level. Development readings carry no band and are never a basis for a hiring, promotion or appraisal decision (Section 12.8).

Confidence is printed on the credential and shown in the register. An employer can distinguish a High-confidence Certified from a Moderate-confidence Certified.

11 Reliability, standard setting and automated scoring

11.1 **Standard setting.** The thresholds in Section 12 were set by AIR APAC on the review board's advice as initial cuts, from the anchor meanings (Level 2 is the threshold behaviour) and the synthetic panels in checks V1 to V3. They are reviewed each year against the evidence programme (Annex A), including classification consistency near the cut score (Section 11.8). A standard-setting study with practitioners in a role family is run on the first real criterion data for that family (Annex A19) and published with its method. A change to a threshold is a major version.

11.2 **Scoring agreement.** Rule-based scoring by an approved instrument must reproduce: the same transcript always yields the same observations and levels, tested on every release. Where human assessors score, independent, authorised assessors scoring a common transcript set must reach **quadratic-weighted κ ≥ 0.70 per dimension** (Gwet's AC2 substituted where base rates are skewed enough to make κ unstable). A

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

dimension below the threshold has its anchor or its rule rewritten in the next version. The analysis plan, transcript and assessor sampling basis, uncertainty and agreement statistics are published per version.

11.3 **Discriminant separation.** Dimensions within a domain are expected to show an observed disattenuated correlation below 0.85. The evidence programme reports it each year; a pair above the threshold is handled under Section 5.1.

11.4 **Rule-based, automated and model-assisted scoring.** Deterministic rule-based scoring by an approved instrument, applying published rules (Annex F) to the retained trace, generates observations directly: every level cites the events and the rule that produced it, and no human confirmation is needed. A model-generated (for example large-language-model) score is **not** an observation and never stands alone:

- A model may propose a level with cited transcript evidence.
- A qualified human assessor confirms or overrides, with reasons recorded.
- The pair counts as **one** observation, attributed to the human assessor.
- Disagreement of more than one level between model and human routes the observation to a second human assessor, whose decision is final.
- Model-assisted proportion is recorded on the result and disclosed in the register entry.
- Model version and prompt version are pinned in the transcript.

11.5 **Drift.** Model-assisted scoring is re-benchmarked against a held-out human-scored transcript set on each model or prompt change, and at minimum quarterly. Failure to meet Section 11.2 against human scores suspends automated scoring for the affected dimensions.

11.6 **Assessor authorisation.** Where a human assessor scores or confirms an observation, only an assessor authorised under AIR APAC's assessor standard may do so. Where an approved instrument scores by rule, AIR APAC issues the certification decision under Section 12. The assessor standard defines entry competence, training, initial calibration, continuing calibration, monitoring, conflict controls, suspension and removal. Reliability testing under Section 11.2 is conducted only with assessors who have completed the initial calibration.

11.7 **Validation programme.** A version of this Standard and its official instrument is validated when the current validation report shows V1 to V5; V6, V7 and V8 ground the prediction and fairness claims in Section 2.1:

- **V1 No passing by ritual.** Blanket strategies (always accept, always challenge, escalate everything, decide at once, never verify) score below a calibrated reference strategy on every scenario, and none reaches Level 2 on any dimension it games.
- **V2 Known profiles land where they should.** Synthetic candidates and persona panels with declared decision styles reach the levels their styles map to, at a rate declared before the run.
- **V3 Stable across scenarios.** The same candidate profile reaches the same level, or one step from it, across parallel scenarios.
- **V4 Readings agree.** Every case reading behind a rule (which advice is sound, which evidence matters) is rated independently and agreed, with agreement recorded.

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

- **V5 Real assessments.** V2 and V3 are re-run on real candidates as assessments accumulate, and on agreement between the instrument's level and an authorised facilitator's independent reading. The report states the count.
- **V6 Prediction.** A level from three scenarios goes with better outcomes on an unseen fourth, by a margin declared before the run. Later performance at work is check V8.
- **V7 Fairness.** Matched candidates who differ only in gender, age, nationality, first language, market or job level reach the same levels within a declared tolerance; real outcomes are monitored under Section 15.
- **V8 Job performance.** In a study of real people in a named role family, Assessment results go with later rated performance at work, by a margin and on a criterion declared before the run. The criterion is a rating form AIR APAC writes, completed by people blind to the results, with hard outcomes from the employer's records where available (Section 4.5 applies: the employer supplies ratings, never criteria). A study may be concurrent, on people already in the role, or predictive, on applicants whose results are sealed from the hiring decision; the report and every claim resting on it say which. The report states the role family, the sample, the criterion, the correlation with its uncertainty and, where available, the gain over the employer's own selection method. V8 is synthetic-free: it uses real people only. The sample size and the smallest effect worth reporting are declared before the run. Where available, two raters rate each person. Corrections for range restriction and criterion unreliability are reported separately from the observed value. The report is published whether the result is favourable or not. A result for one role family supports no claim for another. The analysis is checked by the external academic reviewer (Section 17) or an independent psychometrician before publication. A person assessed in a sealed predictive phase is told that the result is sealed from the hiring decision, and keeps their report and any credential.

Synthetic candidates are a legitimate source of validation evidence for V1 to V7, as they are in market research, provided their personas carry the demographic and role detail of the people they stand for; the report labels every synthetic result as synthetic. The report is versioned with this Standard and published on airapac.org.

The validation report also carries an annual impartiality statement: the share of certified candidates who trained with AIR APAC, AIR Academy or a licensed trainer before certifying, and any other overlap between training and certification.

11.8 **Measurement precision.** The validation report states, per track and role edition: the agreement of the composite and each domain across parallel forms (check V3, on real candidates once V5 has them, and on synthetic panels until then, labelled as such); the standard error of measurement at the composite cut and at the dimension and domain floors; and classification consistency at the cut. Reproducing the same levels from the same transcript (Section 11.2) shows that scoring is deterministic; these figures show how precise it is. Where a figure rests on synthetic panels only, the report says so.

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

12 Certification decision rules

12.1 **Non-compensatory.** Strength in one domain does not offset weakness in another. Domain scores are not averaged away into a single number for the purpose of the pass decision.

12.2 **Thresholds.** All six must be met. Failure of any one results in Not Certified.

12.2.1 **Aggregation rule.** A dimension score is the arithmetic mean of its valid observations, a domain score is the arithmetic mean of its dimension scores, and the composite is the arithmetic mean of all nine dimension scores. The evidence programme tests it each year for sensitivity and classification consistency near the cut score.

12.2.2 **Critical-review trigger.** A Level 0 observation that matches a critical-review pattern in Annex D renders the affected dimension Insufficient Evidence pending targeted reassessment on an unexposed decision point. It may not be averaged into a certifying dimension score. A single trigger is not an automatic Not Certified decision.

#	Requirement	Rule
1	Dimension floor	No dimension below 2.0
2	Domain floor	Each of S, T, E, P at or above 2.0 (mean of its dimensions)
3	Composite	Overall mean at or above 2.3 (Section 11.1)
4	Evidence	Section 8.3 satisfied for all nine dimensions
5	Stakes coverage	Section 6.4 satisfied
6	Confidence	Moderate or High

12.3 **What each rule does.** Stated so the rules are not mistaken for redundancy. A profile at exactly 2.0 across all nine yields a composite of 2.0 and fails rule 3, so the **composite is the binding constraint on a flat profile**. The dimension and domain floors do different work: they stop a candidate clearing 2.3 on the strength of three high dimensions while sitting at 1.0 on, say, E1. Rule 5 stops a candidate clearing on three financially-framed scenarios. Rules 1–3 govern level; rules 4–6 govern whether the level is knowable.

12.4 **Bands.** Certified results are reported in bands, not raw composites, to prevent false precision and league-table misuse.

Band	Composite	Additional condition
Certified	2.3 – 2.9	All floors met
Certified with Merit	3.0 – 3.4	No dimension below 2.5
Certified with Distinction	3.5 and above	No dimension below 3.0

12.5 **Reporting to the candidate.** Every candidate receives: every separately reportable dimension score, the four domain scores, the band, the confidence level, the stakes profiles assessed, the anchors applied, the

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

model-assisted proportion, and the evidence cited for each observation. This is provided whether or not the candidate certifies. Every report and credential carries the Section 2.2 scope statement in plain language.

12.6 **Reporting to employers.** An employment-grade result (Section 12.8) is AIR APAC's expert reading of the person's judgment in AI-enabled work, which the employer may rely on as one input. AIR APAC is responsible for applying this Standard correctly and for the reading it gives. It does not make, and does not accept responsibility for, the employer's decision about the person. The result describes the person's judgment in AI-enabled work at the level certified (Section 2.2); no AIR APAC, partner or licensee material describes it as a finding that a person is fit for a role, job-ready or suitable for a position. Candidates are told before they start who receives their result and what it may be used for. The terms every commissioning organisation accepts are in Section 12.11.

- **Employees.** Where an employer commissions an assessment of its own people, it receives each candidate's result by name: dimension and domain scores, band, confidence and the evidence behind each observation. Its facilitator receives every result; each line manager receives the results of their own people. Before an Assessment, candidates are told an employment-grade result may be used for development, assessment, promotion and hiring. Before a session, they are told it gives a development reading, and that a hiring or certification decision uses a separate Assessment. Development readings reach the employer for development only. Participants are told, before they start, which roles receive a development reading by name (for example the facilitator and their own line manager). The employer is never told whether an access support was used.
- **Applicants.** Where an employer, or a recruiter or staffing firm acting for it, commissions an assessment of people applying for a role, it receives each applicant's employment-grade result by name, in the form of Section 12.5, for that hiring decision. Where a staffing or placement firm commissions the assessment, a result passes to a client only where the applicant notice (Section 12.9) named that client, and only to the people deciding on that placement. On the employer's copy, a Not Certified outcome is labelled 'Did not meet the Standard on this Assessment' with its confidence; the register records Not Certified. Where an applicant's result is Provisional or Insufficient, the employer is told only that no employment-grade result was obtained, and AIR APAC offers the applicant one further sitting on an unseen form at no charge to the applicant. The people deciding on the role receive the result; nobody else in the organisation does. The employer is never told whether an access support was used. AIR APAC provides no rank order of applicants; the band is the reading (Section 12.4). The applicant receives their own report under Section 12.5 no later than the employer receives its copy, with the notice in Section 12.9.
- **Independent candidates and students.** Where a candidate presents independently, the result is released only to parties they name. Campus results follow Section 13.4.

12.7 **No public ranking.** AIR APAC publishes no individual league tables or employer-comparative candidate scores. The commissioning organisation receives a team view: counts of levels and bands across the people it assessed, for the use of people who are not entitled to named results. A team view covers ten people or more; below ten, no team view is given, and the organisation has only the named results it is entitled to under Section 12.6. For development use, a team view counts decisions and how the published rules read them, not levels or bands (Section 20.3). AIR APAC publishes outcome rates by track and by market in the annual summary (Section 15.4), and by institution or employer only in aggregate and never as a comparison.

12.8 **Employment-grade results.** Only an Assessment result from an approved instrument, at Moderate or High confidence, is an employment-grade result. A valid credential shown by its holder is an employment-grade

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

result until its expiry (Section 18.3). Only an employment-grade result may be used as AIR APAC's reading in a hiring, promotion or appraisal decision, whether the outcome is Certified or Not Certified. Practice, a training session, a public demonstration, a Provisional result and an Insufficient result are development readings (Sections 10.1, 20.3): they inform development and decide nothing about a person's employment or certification. AIR APAC releases no development reading to an employer in a hiring context, and every commissioning licence (Section 12.11) binds the organisation not to use a development reading in a hiring, promotion or appraisal decision. A result from a sealed validation study (check V8, Annex A19) is not released for the hiring decision it was sealed from.

12.8.1 **Near-cut results.** Before an employment-grade result is released for a hiring or promotion decision, AIR APAC checks whether a change of one level in any single observation would change the outcome between Certified and Not Certified. Where it would, AIR APAC offers the candidate one further held-back mission on an unseen form at no charge before release, and computes the result on all missions. Once the standard error of measurement is published under Section 11.8, the margin may be set from it instead. The decision rules in Section 12.2 are unchanged; this rule adds evidence where one observation decides the outcome, as Section 12.2.2 already does. The released result states whether the rule was applied.

12.9 **Applicant notice.** Before an applicant starts, they are told in plain language: what the Assessment measures and does not measure (Section 2.2), and how to practise; the employer and the role, and, where a staffing or placement firm commissions the assessment, the firm and any client who may receive the result; that the result goes by name to the people deciding on that role; that it is one input, that no hiring decision is made on this result alone, and that the decision is the employer's; that they may decline, and that if they decline the employer decides without this result; that they receive their own report; what is kept, for how long and why (Section 14.6), that data without their name may be used to improve this Standard and the instrument, that data is held by AIR APAC and Exeter Labs, may be processed outside their country, and is not sold; the contact for access and correction requests (Section 14.5); how to appeal and the time limits (Section 18.5); and how to ask for an access support (Section 8.9). An applicant who is not hired keeps their result and, if they certified, their credential (Section 14.4). Hiring use opens in a market only once the access-support procedure (Section 8.9) and this notice are published for that market (Annex A15, A21).

12.10 **Job relevance.** Before results are used for hiring into a role, AIR APAC records a job-relevance statement: the role family, the decisions in the role that are made with or about AI output, the consequence families they carry (Section 6.1), the track and, where used, the role edition (Section 13.3). AIR APAC writes it from the employer's role description and keeps it with the results. The employer describes the role; it does not set, weight or add criteria (Section 4.5). Where the statement finds that judgment in AI-enabled work is not material to the role, AIR APAC declines the hiring use. The employer's role owner signs the role description the statement is written from. For a new role family, AIR APAC has the statement reviewed by at least one practitioner in that family who does not work for the commissioning employer. The statement records why this Standard is relevant to the role. It is not a job analysis of the whole role and does not say that a person is fit for the role as a whole (Section 2.2). It is disclosed to the applicant or a regulator on request.

12.11 **Commissioning terms.** Every licence under which an organisation commissions assessments or sessions contains these terms, and this Standard relies on them:

- the organisation uses an employment-grade result together with at least one method of its own, and no hiring, promotion or appraisal decision rests on this result alone;

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

- a person reviews any decision in which the result is used;
- the organisation uses no development reading in a hiring, promotion or appraisal decision;
- the organisation has its own lawful basis and gives its own notice for the assessment, and is responsible for its own decision;
- the organisation is asked not to treat a decline under Section 12.9 as adverse;
- where an applicant lodges an appeal within five working days of release, the organisation does not communicate an adverse decision resting on the result until the appeal is decided (Section 18.5);
- the organisation monitors its own selection decisions; AIR APAC monitors its own results (Section 15.7);
- the organisation states, and participants are told before they start, who receives development readings by name (Section 12.6).

13 Tracks

The Standard is single. Tracks differ in scenario complexity, stakes density and stakeholder count. They do **not** differ in dimensions, anchors, thresholds or decision rules.

Track	Population	Distinguishing feature
Campus	Final-year students of institutes of higher learning, and graduates in their first two years of work	Workplace scenarios at entry level: bounded authority with a named supervisor to escalate to, lower stakeholder density, bounded consequence
Professional	Practising managers and individual contributors	Full stakeholder density, organisational consequence, live authority boundaries
Practitioner	Those who lead AI adoption or assess others	Adds multi-party conflict, contested mandate, and ambiguity about who owns the decision
Vendor	Organisations, not individuals	Separate standard (see Section 16.2)

13.1 A Campus certification and a Professional certification are not equivalent, are not convertible, and are recorded distinctly in the register. Any representation of one as the other is misuse. To an employer, a Campus certification means the holder met every certification rule in Section 12 on decisions of the kind an entry-level employee makes with AI in the work. A Campus holder reaches Professional by a full Professional Assessment; the Campus band confers no credit (Section 18.2). That sentence is printed on the Campus credential and carried in the register's scope description (Section 14.1).

13.2 Track is printed on the credential and displayed in the register. A band without a track has no meaning. A candidate eligible for more than one track chooses the track; the register shows the track sat. An employer commissioning the assessment of its own employees uses the Professional track or a role edition of it. An employer commissioning the assessment of graduate applicants may use the Campus track.

13.3 **Role editions.** A track may have role editions, whose scenarios are set in the work of one role family (for example software delivery, finance operations or public service). A role edition uses the same dimensions, anchors, thresholds and decision rules; only the scenario content differs, under Section 8.5 and Annex F. A

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

result from a role edition names the edition on the report, the credential and the register entry. A role edition is approved by AIR APAC before certification use and has its own parallel forms (Section 9.4).

13.4 **Institutions and students.** Where an institute of higher learning commissions Campus assessments:

- The credential belongs to the student (Section 14.4), who has access to their own result (Section 14.5). As a stated term of the commissioned assessment, told to students before they start, the institution receives each student's result by name for teaching and advising, and counts under Section 12.7 for its own cohort reporting, and for nothing else. A Not Certified, Provisional or Insufficient result is held by the student's named adviser, kept apart from the academic record, and is not used for selection, ranking, scholarship, placement, internship allocation or discipline.
- The institution passes no named result to an employer. A student releases their result or credential to employers themselves, by credential code or by naming the party (Section 12.6).
- A Campus result is employment-grade under Section 12.8 when the student releases it to an employer, or when an employer commissions a Campus-track Assessment of its own graduate applicants under Sections 12.6 and 12.9 to 12.11.
- A Campus result is never a condition of graduation. It may form part of a course grade only where the institution published the rule to students before they enrolled in the course, offers an alternative route to the same grade, and never fails a student on this result alone. Practice and teaching inside a course may be graded as the institution sees fit.
- A student may decline the Assessment with no academic penalty, except where a published course rule under the previous point applies.
- Practice and teaching inside a course are development (Section 20). Campus certification needs a separate Assessment on held-back scenarios (Section 9.3).
- The institution is responsible for its own lawful basis and notice to students; AIR APAC provides a standard student notice (Annex A21).
- Fairness monitoring reports Campus outcomes by institution, institution type and programme discipline (Section 15.2), to the institution and in aggregate. AIR APAC publishes no institution's rate as a comparison, and an institution does not publish its rate as a comparison with another institution.
- In prospectuses and course pages, an institution describes the credential in the words of Sections 2.1 and 13.1 only.

13.5 **Campus capacity.** Before an institution commissions a cohort, AIR APAC states in writing: the certifying forms available to that cohort (Section 9.4, Annex A20); the sitting window; the retakes allowed and the form each uses; the identity-check method and who staffs it; and the exposure threshold at which a form is withdrawn (Section 9.3). Where a form is withdrawn for leakage during a sitting window, students already certified keep their credential, and AIR APAC re-sits affected students at its own cost. A course aligned under Section 20.1 may sit inside a credit-bearing module.

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

14 Register, credential portability and data

14.1 **Register contents.** The register records: credential code, track, role edition where used (Section 13.3), band, confidence, standard version, issue date, expiry date, status, and the plain-language scope description required by Section 2.2. It does not record employer, dimension scores, stakes profiles, or any transcript content.

14.2 **Name publication is opt-in.** A holder's name appears in the searchable register only with their explicit, separately-given, revocable consent. The default is verification by credential code: any party holding the code can verify the entry. Withdrawal of name consent does not affect validity.

14.3 **Verification is free** and requires no account, no registration and no commercial relationship with AIR APAC.

14.4 **The credential belongs to the individual.** It survives change of employment. No employer, sponsor or delivery partner may withhold, revoke, condition, or claim ownership of it, including where they paid for it. This is a term of every delivery-partner contract. The credential is the holder's. The transcript and scoring record are AIR APAC's record of the assessment, which the holder may obtain under Section 14.5.

14.5 **Candidate data rights.** A candidate may obtain their own transcript and result, and request correction of factual errors in it.

14.6 **Retention.** Transcripts and results are retained for as long as AIR APAC and its instrument provider need them for certification, recertification, appeal, the evidence programme and the commissioning organisation's licence. The retention period for each purpose is published and stated in the notices under Sections 12.9 and 13.4. Where a person's transcript or result is kept only for the evidence programme and improvement, it is kept without the person's name and contact details unless they have agreed to a named record.

14.7 **Use of data.** Assessment data is used by AIR APAC and its instrument provider to operate, validate and improve this Standard and the official instrument. It is not sold.

15 Fairness monitoring

15.1 AIR APAC monitors certification outcomes for differential performance across groups, using demographic data that is lawfully collected and never used in scoring. Demographic data is collected as optional, separately from scoring, and a person may decline without effect. The groups collected are set per market with the legal review in Annex A5.

15.2 Monitoring covers, at minimum: outcome rate by track, by role edition, by delivery partner, by market, by language of delivery, by assessment modality, by use (development, employee assessment, hiring, Campus), and, for Campus, by institution, institution type and programme discipline.

15.3 **Differential functioning.** Decision points and observations are the primary units screened for differential functioning where the data support that analysis. AIR APAC may use scenario-level or dimension-level analysis with a published justification. The analysis plan specifies methods suitable for ordinal data and

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

available subgroup sizes, flag thresholds, uncertainty, action rules and the limits of inference. A component showing group-differential performance not explained by the construct is withdrawn pending review.

15.4 **Publication.** A fairness summary is published annually. Where a differential is found, it is published together with the action taken. Absence of a published summary in any year is itself a finding.

15.5 **Bias-free.** The claim rests on validation check V7 (Section 11.7) and on this section's monitoring, published and acted on.

15.6 **Accessibility monitoring.** AIR APAC monitors access-support requests, decisions, use, technical failures and outcomes. This information is separated from scoring and reported only in aggregate. A differential associated with an accommodation, instrument or delivery process is investigated under Sections 4.8 and 15.4.

15.7 **Adverse impact in hiring.** Where results are used for hiring, AIR APAC compares Certified rates, rates of not meeting the Standard, and dimension levels across groups of real applicants, by gender, age, nationality and first language at minimum, for each commissioning organisation and role edition once thirty applicants have sat, and across all hiring use each quarter. A group is flagged where its Certified rate is below four fifths of the highest group's rate and an exact test shows the difference is unlikely to be chance, or where the ratio is below four fifths in two consecutive quarters with at least five people in the smaller group. A flag is a screen, not a finding of discrimination. A flag is investigated under Section 15.3 and reported to the commissioning organisation with the action taken. A component flagged in two consecutive quarters with no explained cause is withdrawn from employment-grade use for that role edition (Section 15.3). Employers monitor their own selection decisions; AIR APAC monitors its own results. Synthetic pairs under check V7 do not replace this check.

16 Relationship to other AIR APAC instruments

16.1 **AI Readiness Index.** Rates organisations on public signals. Different unit of analysis, different signal model, different evidence base. An organisation's Index position is **not** a function of its employees' certifications, and a certification confers no Index effect. The two must not be co-marketed in a way that implies conversion between them.

16.2 **Vendor Certification.** Assesses products and providers against a separate standard. Cross-reference to be added when that standard is published. A vendor certification says nothing about the judgment of any individual using the product.

16.3 **The official instrument. StepSim is the official instrument of AIRS-J/1.** It is an AIR APAC initiative, built and operated by Exeter Labs (Section 4.9), and its specification is incorporated in Annex F. Observations StepSim generates under that specification count as AIRS-J/1 evidence. Further instruments may be approved against Sections 8, 9 and 10, and AIR APAC publishes the list of approved instruments and their owners. The rubric in Section 7 depends on no product's interaction design: an instrument implements it, it does not define it.

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

17 Governance roster

Role	Holder
Standard owner	AIR APAC Pte. Ltd. (UEN 202600913G)
External academic reviewer	Prof. Mudiarasan Kuppusamy, UNITAR International University
Review board	Drawn from AIR APAC's standing councils, published at airapac.org/about. The owner of AIR APAC holds the veto (Section 4.2).
Appeals reviewer	A review board member, or an authorised assessor under §11.6, not involved in the original decision or delivery, and not employed by or paid through Exeter Labs

17.1 The board's membership is the councils' published membership. A member with an interest in a matter before the board declares it and does not advise on it.

18 Recertification, suspension and appeal

18.1 **Validity.** Two years from issue.

18.2 **Recertification.** Full reassessment against the then-current version. Prior band confers no credit, no exemption and no reduced evidence requirement.

18.3 **Version transition.** A credential issued under an earlier version remains valid to its expiry and is displayed with its version designation. AIR APAC publishes a plain statement of what changed between versions so a reader can interpret an older credential.

18.4 **Suspension and revocation.** Grounds are limited to: impersonation; use of external AI tools during a certifying session contrary to Section 9.2; possession of leaked scenario content; falsification of identity or eligibility; or material misrepresentation of the credential's meaning after issue.

Process: written notice with the specific allegation and the evidence; 21 days to respond; decision by a panel of three drawn from the review board; written reasons; right of appeal under 18.5. The credential remains valid during the process unless the allegation is impersonation.

18.5 **Appeal.** A candidate may appeal a certification-status decision, any employment-grade result released to an employer (Sections 12.6, 12.8), an accommodation decision that affects access to assessment, or a suspension within 30 days. The grounds in Section 18.6 apply. An appeal on a result released for hiring is decided within ten working days. The applicant receives the result no later than the employer. Where the appeal is lodged within five working days of release, the commissioning organisation holds any adverse communication resting on the result until the decision (Section 12.11). The parties who received the result are sent the decision and, where it changed, the corrected result, which replaces the first for every recipient. Because rule-based scoring is deterministic, re-running the rule is not the review; the review looks for a wrongly applied rule or anchor, evidence the instrument extracted wrongly, a technical failure, or a procedural

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

failure. A development reading decides nothing and is not appealable; a comment on it is a complaint under Section 18.7. Appeals are reviewed by the appeals reviewer under Section 17 who was not involved in the original decision or delivery. The review considers the transcript, published anchors, relevant access-support record and applicable procedures. Outcome and reasoning are provided in writing. Appeal is free for the first instance.

18.6 **Grounds for appeal** are: misapplication of an anchor; an observation not supported by transcript evidence; a procedural failure under Sections 4, 8, 9, 10, 11 or 15; denial or material misapplication of an access support; an undisclosed assessor conflict; a material technical failure; or a localisation or translation defect capable of changing the construct-relevant demand; the near-cut rule in Section 12.8.1 not applied; the sitting mode required by Section 9.1 not met. Disagreement with an anchor itself is not a ground for appeal; it is a comment on the Standard, and is routed to the review board.

18.7 **Complaints.** A complaint is distinct from an appeal and may concern service quality, conduct, a delivery partner, an approved instrument, AIR APAC or a certified person. AIR APAC publishes a complaints procedure covering intake, acknowledgement, independent investigation, response times, escalation and aggregate reporting. A complaint does not change certification status except through the due process in Sections 18.4 to 18.6.

19 Version history and change control

19.1 **Owner.** AIR APAC Pte. Ltd. is the issuing body and owns this document. The owner of AIR APAC approves every version; the review board reviews every major version before release (Section 4.2).

19.2 **Versioning.** Compatible refinements (clarifications, errata, added guidance, added annex material) are released as v2.x. Material changes to the competency model, the anchors, the evidence requirements or the decision rules are released as v3.0. Each released version keeps a permanent URL and is never overwritten; the entry below records what changed and why.

19.3 **Corrections and feedback.** Anyone may submit a correction, an objection to a requirement, or evidence that contradicts a claim made here, through the contact route published at airapac.org/standard. Submissions are logged. A submission that identifies a defect in a normative requirement is put to the review board, and the outcome is recorded with the next version. Adverse findings are published under Section 4.8.

19.4 **History.**

Version	Date	Change
0.1	Draft	Initial draft. Nine dimensions proposed. One anchor set written. Not issued.
0.2	2026-08-15	Reconciled with canonical StepSim scenario methodology. Added E1 Decision Rights and Escalation; merged stakeholder alignment and disclosure into E2; folded proportionality into P1. Added stakes profile and coverage requirement (§6). All nine anchor sets written. Level 4 rewritten to remove positional-authority confound. Added non-claims (§2.1), assessment integrity (§9), reliability and standard setting (§11), fairness monitoring (§15), candidate data rights (§14.5–14.7). Corrected Merit/Distinction band ordering. Not issued.

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

Version	Date	Change
0.3	2026-08-15	External-scrutiny revision. Clarified certified function and managerial scope; removed premature instrument approval; added access-support rules, assessor authorisation, explicit provisional aggregation, critical-review reassessment, confidence precedence, performance-appropriate standard-setting requirements, pilot discriminant gating, expanded fairness analysis, appeals and complaints. Expanded the v1.0 evidence gates. The nine dimensions and anchors are unchanged. Not issued.
1.0	2026-08-19	Initial Edition. Published as the authoritative specification. No change to the competency model, anchors, evidence requirements, decision rules or non-claims: v1.0 is v0.3 text with the publication decision made and the gating vocabulary corrected. The requirements that were written as gates on "v1.0 issue" are gates on opening certification, and are restated as such throughout, including Annex A. Added the status-of-this-edition note, the owner, versioning and corrections provisions (§19.1 to §19.3), and an effective date.
1.0-E1	2026-08-23	**Erratum to v1.0**, issued under the interim-errata provision in the header. Annex B.1 stated that every dimension maps to exactly one lifecycle code and that no dimension spans two. That contradicted the table it introduced, in which T3, E2 and P1 each span two, and it located duplicate prevention in the mapping rather than in the evidence. The statement is withdrawn and replaced: the crosswalk is many-to-many, the three spanning dimensions are marked in the table, and duplicate prevention is restated as the evidence-level rule that no single piece of participant evidence may be scored under more than one dimension. **No normative change.** No dimension definition, anchor, threshold, evidence requirement, decision rule or non-claim is altered.
1.1	2026-10-01	**Compatible refinement** under §19.2, issued without the review board because §4.2 requires the board only before a major version. StepSim is no longer named as a candidate instrument (§16.3, A12); instrument neutrality is unchanged. Added §4.9, the issuing body's declared interests. Annex B.1 and B.2 no longer route to StepSim's methodology: the lifecycle is AIR APAC's scenario-authoring checklist, and Section 6 is the canonical stakes-profile definition. **No normative change** to any dimension, anchor, threshold, evidence requirement, decision rule, non-claim or Annex A gate.
2.0	2026-10-03	**Major version**, approved by the owner of AIR APAC and by the review board drawn from AIR APAC's councils (§4.2, §17) on 2026-10-03; no dissent recorded. AIRS-J/1 becomes an operating certification and AIR APAC the judgment authority. StepSim is the official instrument (§8.1, §16.3, Annex F). Non-claims (§2.1) become claims with named bases; the validation programme (§11.7) is added, including synthetic-candidate evidence, prediction (V6) and fairness (V7). Rule-based instrument scoring generates observations directly; model-generated scores still never stand alone (§11.4). A response bank of authored lines with decoys is a response format, not prompting (§7.1.2). C4 includes wrong or misused records about identifiable people (§6.1). Several scenarios in one sitting are not clustering for confidence (§10). Results go by name to the commissioning organisation, with candidates told before they start (§12.6). The review board is drawn from AIR APAC's standing councils, with the owner's veto (§4.2, §17). Retention follows need (§14.6). Annex A becomes the evidence programme and no longer blocks certification. High confidence no longer needs two instrument types (§10). **Unchanged:** the nine dimensions, the anchors, the 0 to 4 scale, the stakes profile and coverage rule, the minimum observations, the thresholds and the bands.

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

Version	Date	Change
2.1	2026-10-15	**Compatible refinement** under §19.2, approved by the owner of AIR APAC; the review board told (§4.2). Aligning the Standard to three uses on one rubric: development (training), readiness (Campus) and assessment for employment (§2). Title becomes "Standard for Judgment in AI-Enabled Work"; designation and the meaning of "managerial" (§2.2) unchanged. "Predictive of job performance" rests on a new criterion check on real people, V8, by role family; V6 supports it and is not its basis alone (§2.1, §11.7), as the owner approved on 4 October 2026. Employment-grade results defined (§12.8); applicants distinguished from employees (§12.6), with a notice (§12.9), a job-relevance statement (§12.10), an adverse-impact check (§15.7) and appeal (§18.5). The owner takes no part in any individual result (§4.3). Campus covers final-year students and recent graduates on entry-level workplace scenarios, with rules for institutions (§13, §13.4). Role editions (§13.3). Team view floor of ten (§12.7). Training aligned to the Standard (§20). Facilitator certification declared in §4.9. Evidence programme items A19 to A21. Who receives what changes: an applicant's Provisional or Insufficient result no longer reaches the employer (v2.0 §12.6 sent every commissioned result by name), and an employment-grade result that one observation would flip gains one further mission before release (§12.8.1), by the evidence-adding route already in §12.2.2. New obligations on AIR APAC and on users of results; **no change** to the competency model (§5), the anchors (§7), the minimum observations (§8.3), or the certification thresholds, bands and decision rules (§§12.1 to 12.5).

20 Training aligned to the Standard

Training is the development use of this Standard (Section 2). This section says what a training programme may claim. It does not set what a programme teaches or how it is delivered.

20.1 **"Aligned to AIRS-J/1".** A programme, including one delivered by a licensed trainer or a delivery partner, may describe itself as aligned to AIRS-J/1 only where:

- each learning outcome names the dimension or dimensions it develops, in the words of Section 5;
- its practice scenarios meet the eligibility rule in Section 6.3, and together cover at least two consequence families, one of them C1, C4, C5b or C6;
- feedback to participants refers to the anchors in Section 7.5, and no criterion of the trainer's own is presented as part of the Standard (Section 4.5);
- it uses no held-back Assessment content and no leaked content (Sections 9.3, 9.5).

A programme states its outcome map on its course page or in its funding application. AIR APAC may ask for it, lists aligned programmes on request, and removes a listing where this section is not met. A programme that does not meet this section may say it is informed by AIRS-J/1 and makes no alignment claim. Alignment is not an accreditation by any funding body and does not by itself qualify a programme for funding.

20.2 **Completion records.** A record of completion states what the person attended and practised: participant, programme, date and hours, the dimensions practised, the number of scenarios played and their consequence families, the sentence 'This is not an AIRS-J/1 credential', and where to book an Assessment. It carries no level, band or confidence, is not an AIRS-J/1 credential, and confers nothing toward certification. A trainer may issue the attendance or completion document a funding body requires, provided it does not

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

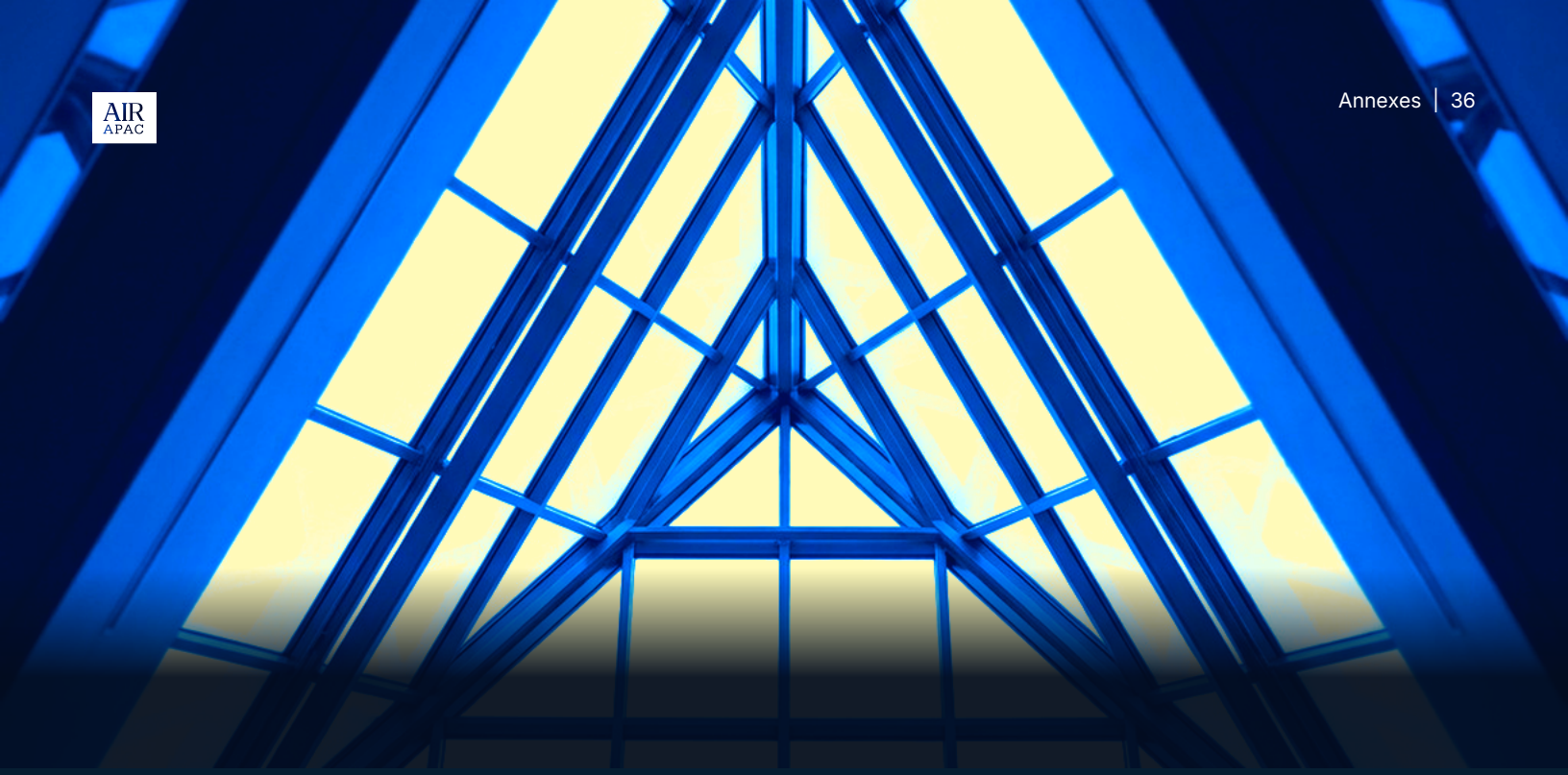
present the AIRS-J/1 designation as a credential, level or band and does not imply certification. A trainer may tell participants that certification is open to them and where to book it, and may not say or imply that attendance raises the chance of certifying, that a participant is ready, or that a result is assured (Section 4.4). Certification needs a separate Assessment on held-back scenarios (Sections 8.4, 9.3).

20.3 **Development readings.** A reading from practice or a training session shows the person's decisions and how the published rules read them, labelled as a development reading. Levels per dimension and bands come only from an Assessment. Who sees it is stated to participants before they start. It decides nothing about a person's employment or certification (Section 12.8).

20.4 **Progress claims.** A programme may claim that participants' judgment improved only on this basis: before and after the programme, each participant plays unseen parallel practice scenarios of the same track and comparable stakes coverage. The claim is made for a group of at least ten participants, never for a named person. It states the number of participants, the scenarios, and the share of decision points at which the published anchors read the participants' acts higher after than before, by dimension where the scenario elicits it. It gives no level, band or confidence, because those come only from an Assessment (Section 20.3). Improvement on a scenario the person has already played is practice, not progress. A claim of improved judgment is not a claim of readiness for certification.

20.5 **Trainers.** A trainer who delivers an aligned programme need not hold an AIRS-J/1 certification. Facilitator certification is a delivery credential issued by Exeter Labs (Section 4.9); it is not a certification under this Standard. A trainer who scores or confirms an observation for certification acts as an assessor under Sections 4.6 and 11.6. Section 4.4 applies to every trainer.

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.



Annexes

A Evidence programme 37

B Crosswalk and retirement 39

C Assessor observation record 41

D Critical-review patterns 42

E Informative global benchmarks 43

F The official instrument 44

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

ANNEX A

Evidence programme

These items run alongside certification from v2.0 and strengthen it. None blocks certification. Progress is recorded in the validation report (Section 11.7).

#	Item	Owner
A1	Review board drawn from the standing councils (Section 17); its members told, and v2.0 reviewed, before release.	AIR APAC
A2	Scoring agreement: the official instrument's reproducibility test on every release; human-assessor agreement where humans score (§11.2).	AIR APAC
A3	Annual review of the cut score and bands against classification consistency (§11.1).	Review board
A4	Observations per scenario and total assessment time per track, published from real assessments (§8.4).	AIR APAC
A5	Legal review of §§12.6, 12.9, 12.11, 13.4, 14 and 15 per market where certification is offered, including the demographic groups that may be collected (§15.1), cross-border processing, and whether separate consent or an impact assessment is required for transfer.	AIR APAC
A6	Parallel forms and the exposure-retirement threshold (§§9.3, 9.4): held-back Assessment scenarios for the official instrument.	AIR APAC
A9	Discriminant analysis on real assessments (§§5.1, 11.3).	Review board
A10	Vendor Certification cross-reference (§16.2).	AIR APAC
A11	Seeding sequence for the public register.	AIR APAC
A12	Approval criteria for further instruments beyond the official instrument (§8.1).	AIR APAC
A13	Job and practice analysis linking the dimensions and anchors to documented workplace evidence.	Review board
A14	The validation programme (§11.7), V1 to V8, reported each version.	AIR APAC
A15	Accessibility and accommodations procedure (§§7.1.1, 8.9), published per market before hiring use opens there (§12.9).	AIR APAC
A16	Assessor standard and complaints procedure, for human-scored instruments and appeals (§§11.6, 18.7).	AIR APAC
A17	Translation and adaptation procedures, before certification in each localised language. Blocks only that language version.	Review board
A18	Annex D critical-review patterns, confirmed or amended from real assessments.	Review board

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

#	Item	Owner
A19	First V8 criterion study, with one employer or recruiter as criterion partner: concurrent phase on people already placed, then predictive phase with results sealed (§11.7). A second partner in the same role family for replication. The standard-setting study under §11.1 runs on the same data.	AIR APAC
A20	Campus forms: two parallel forms of entry-level workplace scenarios, the first institution partner, and an informative Campus implementation note (sitting length, practice-to-Assessment ratio, and mappings to the national skills and qualifications frameworks in Singapore and Malaysia) (§§9.4, 13, 13.4, 13.5).	AIR APAC
A21	Templates for the job-relevance statement, the applicant notice and the student notice (§§12.9, 12.10, 13.4), with a legal reading per market where hiring or Campus use is offered (with A5).	AIR APAC

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

ANNEX B

Crosswalk and retirement

B.1 AIRS-J/1 dimensions to the scenario-authoring lifecycle

The scenario-authoring lifecycle U1–U5 is retained as an **authoring and elicitation checklist**. It is not a scoring model, is not reported, and does not appear on any credential.

The crosswalk is many-to-many, and deliberately so. One lifecycle stage elicits several dimensions: U1 elicits both S1 and S2, U2 elicits T1, T2 and part of T3, and U5 elicits P2 plus parts of E2 and P1. Three dimensions draw on two lifecycle stages each, marked below. Read the table in one direction only: it says where in a scenario each dimension's evidence is elicited. It does not assign a scenario stage to a dimension, and a lifecycle code is never a proxy for a dimension score.

AIRS-J/1 dimension	Lifecycle stage	Spans two	Origin construct
S1 Problem Framing	U1		Decision framing; nondelegable responsibility
S2 Scope and Constraint Definition	U1		Decision framing; material unknowns
T1 Evidence Verification	U2		Evidence and AI calibration (U2a, U2b)
T2 Assumption and Alternative Testing	U2		Competing explanations and discriminating action; falsifiers (U2c)
T3 Risk and Failure Anticipation	U2 + U3	yes	Failure modes, cost bearers and accepted harms; uncertainty (U2c) carried into trade-offs (U3)
E1 Decision Rights and Escalation	U4		Decision authority (U4a); escalation and execution ownership (U4c)
E2 Stakeholder Alignment and Disclosure	U4 + U5	yes	Data, disclosure and action boundaries (U4b); audience-appropriate communication (U5b)
P1 Proportionate Commitment Under Uncertainty	U3 + U5	yes	Trade-offs and proportionality (U3); commitment and handoff (U5a)
P2 Consequence Ownership and Correction	U5		Revision and adaptation (U5c)

What actually prevents duplicate scoring. Duplicate prevention is an evidence-level rule, not a mapping-level rule. The binding constraint is that **no single piece of participant evidence may be scored under more than one dimension**. Two dimensions elicited at the same lifecycle stage are permitted, and are common; two dimensions scored from the same sentence, action, or artifact field are not. Where an assessor finds only one piece of evidence spanning what look like two dimensions, they score the dimension the evidence most directly demonstrates and record the other as not evidenced, rather than scoring both. Section 8 governs evidence sufficiency; this Annex does not create an exception to it.

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

An earlier revision of this Annex stated that every dimension maps to exactly one lifecycle code and that no dimension spans two. That was inconsistent with the table it introduced, in which T3, E2 and P1 each span two, and it misidentified where duplicate prevention comes from. The statement is withdrawn and replaced by the two paragraphs above. No dimension definition, rubric anchor, threshold, or decision rule changes as a result.

Why P1 merges proportionality and commitment. Both are evidenced by the same act: the posture chosen and its stated rationale. Scoring them separately would score one sentence twice without distinct rubric evidence, which the existing duplicate-prevention rules prohibit. They are merged rather than dropped.

Why E1 is new. Decision authority and nondelegable responsibility are elicited by every current scenario and were not represented in the v0.1 nine. Their absence would have made the model unusable for the enterprise cases it is aimed at.

B.2 Consequence and qualifier codes

C1–C6 and the four contextual qualifiers are adopted unchanged as the stakes profile (Section 6). They remain unscored. From v1.1, Section 6 of this Standard is their single canonical definition.

B.3 Retired: the ten-domain readiness model

The domain set **Context Engineering, AI Direction, Discernment, Trust Calibration, Communication, Adaptive Effectiveness, Governance and Disclosure, Ownership and Defense** and associated constructs is **retired**. It was recorded as a design hypothesis, never had a factor structure, and cannot coexist with AIRS-J/1 without destroying comparability.

Indicative mapping, for reading historical material only. It is not a conversion and no historical score converts to an AIRS-J/1 result.

Retired domain	Nearest AIRS-J/1 dimension
Context Engineering	S1, S2
AI Direction	T1 (partial); otherwise out of scope; tool proficiency is not assessed
Discernment	T1, T2
Trust Calibration	T1, P1
Communication	E2
Adaptive Effectiveness	P2
Governance and Disclosure	E1, E2
Ownership and Defense	P2

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

ANNEX C

Assessor observation record

Minimum fields for a single observation. An observation missing any field is void.

Field	Content
Observation ID	Unique within transcript
Candidate credential code	Linked to the candidate's identity in the commissioning organisation's records (§12.6), or in AIR APAC's records where the candidate presents independently or through an institution (§13.4)
Dimension	One of nine
Decision point ID	Must be distinct from other observations of the same dimension
Level awarded	0–4
Anchor text applied	Verbatim from Section 7.5
Transcript evidence	Direct citation of the candidate's action, artifact or statement
Scaffolding level	None / prompted / guided; caps level per §7.1
Access support	None / approved support code; does not cap level under §7.1.1
Critical-review trigger	None / Annex D pattern code / pending or resolved status
Stakes profile	Primary and secondary consequence family
Scoring mode	Rule-based instrument / human / model-assisted-human-confirmed
Model + prompt version	Where model-assisted
Assessor ID	Human assessor of record, or the instrument and rule version for rule-based scoring
Conflict declaration	Per §4.6, where a human assessor scores
Second scoring	Where applicable: second assessor ID, level, resolution
Timestamp	Observation recorded

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

ANNEX D

Critical-review patterns

This annex supplies no automatic failure rule. A narrow set of Level 0 observations interrupts aggregation and requires targeted reassessment under Section 12.2.2. The review board confirms or amends the patterns from real assessments (Annex A18).

A pattern is triggered only by transcript-observable behaviour. An assessor may not infer dishonesty, motive or intent. The relevant stakes pathway and the evidence supporting the trigger must be cited.

Code	Dimension	Pattern
CR-T1	T1	Represents a load-bearing claim or source as verified when the retained trace shows it was not verified, verification was available within the provided assessment environment before commitment, and action on the claim creates a credible irreversible or non-financial harm pathway.
CR-E1	E1	Commits an action outside an authority boundary explicitly available in the assessment, where the action creates a credible irreversible or non-financial harm pathway.
CR-E2	E2	Omits material AI involvement, uncertainty, limitation or self-interest from a party explicitly entitled to know, where the omission could materially change that party's action.
CR-P2	P2	After new evidence disproves the committed position, preserves the action while transferring responsibility to the AI system, data or another party, where the resulting harm pathway remains active.

The evidence programme records trigger frequency, subgroup distribution, targeted-reassessment outcome and effect on classification. A pattern that cannot be applied reliably and fairly is removed or rewritten.

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

ANNEX E

Informative global benchmarks

These references inform design and external review. They are not incorporated as undisclosed scoring criteria, and their inclusion does not assert conformity, accreditation or legal compliance.

- ISO/IEC 17024:2026, **Conformity assessment: General requirements for bodies operating certification of persons.**
- ISO 10667-1:2020 and ISO 10667-2:2020, **Assessment service delivery: Procedures and methods to assess people in work and organisational settings.**
- AERA, APA and NCME, **Standards for Educational and Psychological Testing** (2014).
- International Test Commission and Association of Test Publishers, **Guidelines for Technology-Based Assessment** (updated 2025).
- NIST, **Artificial Intelligence Risk Management Framework 1.0** (2023).
- ISO/IEC 42001:2023, **Information technology: Artificial intelligence: Management system.**

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

ANNEX F

The official instrument

StepSim's instrument specification, "AIRS-J/1 instrument specification: StepSim" version 2.0.0, is incorporated in this Standard by reference and is published with it on release. It sets out how StepSim's recorded acts become observations for each of the nine dimensions, the rule for each level, each scenario's decision points and stakes profile, the number of Assessment scenarios a certifying result needs, and how StepSim meets Sections 8, 9 and 10. A change to the specification that changes a rule is released with a new version of this Standard.

Ends. AIRS-J/1 v2.1, approved 2026-10-15. Certification is open, with StepSim as the official instrument.

AIRS-J/1 v2.1 · © 2026 AIR APAC. All rights reserved.

AIR APAC Center for AI Readiness

Published in full. No scoring criterion applied in certification exists outside this document and the official instrument's specification (Annex F).

AIRS-J/1 version 2.1, dated 2026-10-15, is published at airapac.org/standard/v2-1, which carries any errata issued since this PDF was produced on 23 October 2026.

Read online

airapac.org/standard/v2-1

Certification

airapac.org/standard/certification

Verify

airapac.org/standard/register/verify

© 2026 AIR APAC. All rights reserved. AIR APAC Pte. Ltd. (UEN 202600913G), Singapore.