APAS: Autonomous Pentest Assurance Standard

Version: v0.8

Status

This document is the Core Standard for APAS. It defines the minimum normative concepts, quality dimensions, claim structure, evidence requirements, and interpretation rules for autonomous pentest assurance.

Detailed scoring formulas, evaluator run procedures, profile-specific thresholds, benchmark controls, truth-set construction, framework crosswalks, and safety/dangerous-capability evaluation procedures are out of scope for this Core document and SHALL be defined in companion documents. APAS evaluation as defined here assesses the truthfulness and usability of pentest evidence; it does not by itself certify the safety, containment, or misuse resistance of the candidate system, and APAS certification SHALL NOT be represented as a fitness-for-deployment safety assurance.

Normative keywords: MUST, MUST NOT, REQUIRED, SHALL, SHALL NOT, SHOULD, SHOULD NOT, MAY.

Informative note on v0.8

APAS v0.8 preserves the eight-dimension claim model introduced in v0.6 and makes four clarifications explicit:

Theme	v0.8 clarification
Taxonomy posture	APAS is a taxonomy and attestation standard first. Comparative testing MAY be used, but APAS is not a leaderboard or benchmark badge.
Quality model	APAS quantifies autonomous pentest quality through declared quality dimensions: coverage fidelity, finding validity, evidence traceability, remediation usability, retestability, and process integrity.
Coverage discipline	Higher-assurance and compliance-oriented claims require a structured coverage declaration and, where applicable, a negative-result statement.
Anti-proxy rule	Higher execution autonomy, lower human participation, or fewer or no reported findings SHALL NOT by themselves be treated as proxies for better pentest quality.

APAS v0.8 continues to separate execution autonomy, collaboration model, assurance outcome, and reliance outcome. APAS v0.8 also continues to make a second point explicit:

Context principle	Standard position
Rich context preference	context-rich, code-aware testing is preferred for higher-assurance claims
Mode equivalence	black-box and grey-box testing remain valid modes, but SHALL NOT be presumed equal in efficiency, coverage fidelity, signal quality, or evidence quality to white-box or code-aware testing
Profile preference	profiles intended to support high-assurance, assessor-ready, or substitution-oriented outcomes SHOULD prefer the richest lawful context available

1. Purpose

APAS exists to answer one question:

Under what conditions may an autonomous system produce pentest evidence that a relying auditor, assessor, or customer-appointed reviewer may treat as decision-grade within a declared scope?

APAS is intended to:

- define bounded and auditable claims for autonomous pentest capability and quality
- quantify pentest quality through declared, reviewable quality dimensions rather than unsupported marketing language
- separate execution autonomy from human collaboration, assurance outcome, and reliance outcome
- require declared coverage, validated findings, traceable evidence, usable remediation guidance, retestability, and integrity disclosures
- support direct review by a relying party rather than hidden vendor-side interpretation
- support framework-aligned evidence packaging without implying framework acceptance
- provide a common language for buyers, evaluators, assessors, security teams, and compliance programs that interoperates with established vulnerability and assurance standards (CVE, CWE, CVSS, EPSS, SARIF, VEX/CSAF, and OSCAL), so that APAS findings, severities, and attestations can be consumed by existing vulnerability-management and assessment tooling; the binding crosswalks SHALL be defined in the APAS Attestation Schema and Crosswalk Annexes

APAS does not certify that a system is universally equivalent to a human pentester.
APAS is not itself a benchmark program, and public benchmark performance alone does not establish certification or high-assurance claims.
APAS does not itself establish regulatory, contractual, or audit compliance.
APAS certification alone SHALL NOT imply framework acceptance, assessor acceptance, or waiver of framework-specific requirements.

1.1 Core principles

APAS-CORE-1.1.1
All APAS claims SHALL be interpreted only within the declared mode, tier, execution class, collaboration class, assurance outcome class, environment class, affordance profile, reliance outcome, and, where required, the associated quality declaration.

APAS-CORE-1.1.2
Execution autonomy SHALL NOT be treated as a proxy for pentest quality, coverage fidelity, evidence quality, or assessor usability.

APAS-CORE-1.1.3
Comparative benchmark position, corpus performance, or public challenge performance SHALL NOT be treated as a proxy for pentest quality or assessor usability. Benchmark and public-corpus results MAY inform capability calibration and drift detection under the Evaluator Guide and 11.2.3; such use SHALL NOT constitute a certification proxy.

APAS-CORE-1.1.4
No APAS claim SHALL be interpreted beyond its declared scope, stated limitations, declared reliance outcome, or declared quality dimensions.

APAS-CORE-1.1.5
A system MAY have lower execution autonomy and higher assurance outcome than a more autonomous system.

APAS-CORE-1.1.6
Pentest quality under APAS SHALL be expressed primarily through declared coverage fidelity, finding validity, evidence traceability, remediation usability, retestability, and process integrity.

APAS-CORE-1.1.7

Modes with greater implementation context, including white-box and code-aware testing, SHOULD be preferred where the assessment objective is to maximize implementation coverage and reduce noise. Where the assessment objective is to model an external or unauthenticated adversary, black-box testing is the appropriate mode for that objective and SHALL NOT be treated as lower-assurance for that objective.

APAS-CORE-1.1.8

Black-box and grey-box results SHALL NOT be presumed equivalent to white-box or code-aware results absent profile-specific justification or evaluator-validated evidence, under the methodology in force, that the BB or GB result achieved equivalent coverage fidelity, finding validity, and evidence quality for the declared scope.

APAS-CORE-1.1.9
A candidate system MAY support more than one valid APAS capability combination. APAS certifies combinations under declared conditions and SHALL NOT be interpreted as blanket certification of the candidate system outside the attested combinations.

APAS-CORE-1.1.10
Every public or private representation of a multi-configuration candidate system SHALL distinguish between evaluated combinations and unevaluated combinations.

APAS-CORE-1.1.11
A claim for one capability combination SHALL NOT imply any claim for another capability combination solely because the combinations appear adjacent, similar, or operationally related. For any claim represented as APAS-certified at v1.0 or later, the APAS Evaluator Guide, Profile Registry, and Attestation Schema are mandatory; the Core Standard alone is not a certifiable basis. Crosswalk Annexes are required where the applicable profile mandates them (see 4.5.2).

APAS-CORE-1.1.12 No low-finding or no-finding result SHALL by itself be interpreted as complete coverage, high assurance, or substitution readiness. A no-finding result SHALL be accompanied by the negative-result statement required under 4.5.

APAS-CORE-1.1.13
No claim made under richer context, lower collaboration, higher collaboration, lower execution independence, higher execution independence, or a different reliance outcome SHALL be generalized across combinations unless the applicable APAS profile or Evaluator Guide explicitly permits such carry-forward and the evaluator attests that the carry-forward conditions were satisfied.

2. Scope

2.1 In scope

This Core document defines:

- the APAS claim model
- the APAS quality-dimension model
- quality declarations and coverage declarations
- claim syntax and claim semantics
- execution autonomy classes
- collaboration classes
- assurance outcome classes
- reliance outcomes
- core role definitions
- minimum evidence-package requirements
- minimum governance and public-claims requirements
- minimum finding-state and evidence-state semantics

2.2 Out of scope

This Core document does not define in full:

- scoring formulas
- evaluator run procedures
- benchmark design
- truth-set construction
- tier-specific thresholds
- profile-specific coverage formulas
- framework-specific crosswalks
- sector-specific acceptance criteria
- physical, wireless, social-engineering, or unrestricted red-team certification

Such material SHALL be defined, where applicable, in:

- the APAS Evaluator Guide
- the APAS Profile Registry
- the APAS Attestation Schema
- framework-specific annexes or crosswalks

APAS-CORE-2.2.1 No party SHALL represent this Core document alone as sufficient evidence of certification, substitution, or framework alignment without the applicable companion materials. APAS certification applies only to bounded, authorized, scope-limited pentest engagements; no party SHALL represent or operate an APAS-certified system as an unrestricted, unauthorized, or general-purpose autonomous offensive capability, and APAS certification SHALL NOT be cited as endorsement of such use.

3. Terms

For purposes of this Core document:

Affordance Profile
The declared set of capabilities, restrictions, tools, permissions, access conditions, and operating constraints available during testing.

Assessor-Consumable Evidence Package
A complete package consisting of the report, attestation, quality declaration, coverage declaration, evidence manifest, traceability references, exclusions statement, and retest information, structured so that a relying reviewer can evaluate scope, coverage, findings, methodology, limitations, and the claimed reliance outcome without vendor-side analytical interpretation or editing.

Assurance Outcome
The declared class describes the strength, reproducibility, and reviewer usability of the resulting pentest output.

Capability Combination
A unique combination of APAS claim dimensions consisting of mode, tier, execution class, collaboration class, assurance outcome class, environment class, affordance profile, and reliance outcome.

Capability Matrix
A structured representation of the capability combinations a candidate system version supports, has been evaluated for, and may publicly claim under APAS.

Candidate System
The autonomous product, service, or agentic workflow submitted for evaluation or used in an APAS-conformant engagement.

Collaboration Model
The declared class describes how humans and agents interact during the testing workflow.

Combination Status
The declared status of a capability combination within a capability matrix.

Confirmed Finding A finding for which the evaluator or applicable conformance process independently reproduces the vulnerable condition or exploit path, OR, where the finding is time-sensitive, destructive, one-shot, or otherwise not safely reproducible, validates preserved first-capture evidence sufficient for decision-grade reliance within the declared methodology.

Coverage Declaration
A structured representation of the materially tested, partially tested, untested, excluded, unreachable, and context-constrained portions of the declared reachable attack surface, together with the roles, workflows, privilege boundaries, vantage points, and context sources exercised.

Coverage Fidelity
The degree to which the coverage declaration truthfully and specifically represents what was and was not exercised under the declared conditions.

Declared Methodology
The documented methodology, scope model, coverage obligations, quality-dimension requirements, evidence requirements, and reporting expectations applicable to the declared APAS profile.

Environment Class
The declared environment model within which a claim is made.

Evidence Traceability
The degree to which findings, coverage statements, limitations, and retest claims can be traced to supporting artifacts and methodology.

Evaluator
The independent party responsible for determining whether a candidate system or engagement satisfies the applicable APAS requirements.

Execution Autonomy
The declared class describes how independently the candidate system performs the testing workflow.

Finding Validity The degree to which reported findings are validated, confirmed, severity-normalized, and not materially overstated.

Human Assistance
Any human action that materially affects strategy, exploitation, validation, prioritization, control mapping, or report content produced by the candidate system.

In-Band Analytical Human Assistance
Human analytical input that occurs during execution and materially changes strategy, exploitation, validation, prioritization, or report content.

Multi-Configuration Capability Statement
An evaluator-issued or vendor-published statement that references more than one atomic APAS claim for the same candidate system version and presents them in a capability matrix.

Negative-Result Statement
A required explanation for a package with no confirmed findings, or only low-materiality confirmed findings where higher assurance or reliance is claimed, identifying exercised scope, exclusions, untested areas, limiting conditions, and the basis on which the absence of confirmed findings may and may not be interpreted.

Observation
An anomalous signal, artifact, or behavior not yet validated as a finding.

Observed Operating Path
The sequence of materially relevant execution, collaboration, context, and assistance states actually used during a run.

Process Integrity

The truthfulness and integrity of scope compliance, safety compliance, assistance disclosure, collaboration disclosure, and evidence handling. The candidate system SHALL also exhibit agent integrity: resistance to manipulation by adversarial content originating from the target environment (including prompt injection, instruction-override, or jailbreak attempts embedded in web pages, responses, error messages, or files), and the controls that prevent such content from altering the agent's scope adherence, findings, evidence, or out-of-band actions. A claim SHALL disclose the agent-integrity controls exercised and any observed manipulation attempts during the run.

Profile
A normative APAS companion specification that defines additional requirements for a specific testing objective, coverage model, or framework-aligned use case.

Quality Declaration
A structured representation accompanying an APAS claim that records the applicable quality dimensions, coverage declaration reference, finding-state summary, evidence-traceability references, remediation status, retest status, and integrity disclosures for the evaluated combination or engagement.

Reachable Attack Surface The portion of the in-scope surface that the candidate system could lawfully discover or reach under the declared affordance profile, environment class, credentials, vantage points, and context sources. For Q3, Q4, or R3 claims the reachable attack surface SHALL be assessed against an independently-established surface baseline (asset inventory, recon baseline, or prior assessment) and SHALL NOT be defined solely by what the candidate system itself discovered, so that under-discovery cannot be presented as high coverage.

Reference Validation
Comparative validation against a declared reference method, reference cohort, or equivalent baseline under the same declared conditions.

Reliance Outcome
The declared level at which the output may be used by a relying party.

Remediation Usability
The degree to which the package enables practical remediation, root-cause understanding, and prioritization by the relying party or target owner.

Retestability
The degree to which the package and supporting evidence enable later verification of fixes or repeat testing under declared conditions.

Retested FindingA prior confirmed finding that has been evaluated again after claimed remediation, with the retest result explicitly disclosed.

Relying Reviewer
The auditor, assessor, or customer-appointed reviewer who decides whether an APAS evidence package is sufficient for the intended assurance or compliance use.

Unsupported FindingA reported issue that, after adjudication, is either (a) not confirmed under the declared methodology because the condition does not exist or is not exploitable (a false positive), or (b) confirmed in existence but whose reported impact is materially overstated (an over-severity finding). The quality declaration SHALL report these two sub-classes separately and SHALL state the unsupported-finding rate (unsupported findings as a proportion of all reported findings) for the evaluated combination.

Validated FindingA finding for which the candidate system demonstrates exploitability or strong evidence of a vulnerable condition, pending independent confirmation under the declared methodology. A finding MAY be a composed exploit chain whose impact derives from the combination of individually lower-severity conditions; impact substantiation SHALL be assessed against the chain as a whole, not its components in isolation.

Vendor-Side Analytical Review
Any vendor-originated interpretation, editing, prioritization, control mapping, or conclusion-shaping applied after execution and before delivery to the relying reviewer.

4. Quality Model

4.1 Quality dimensions

APAS defines the following core quality dimensions:

Quality dimension	Definition
Coverage Fidelity	the truthfulness and specificity with which tested, partially tested, untested, excluded, and unreachable scope are declared.
Finding Validity	the degree to which reported findings are validated, confirmed, not materially overstated, and accompanied by appropriate impact rationale.
Evidence Traceability	the degree to which findings, coverage statements, limitations, and retest outcomes can be traced to artifacts and declared methodology.
Remediation Usability	the degree to which the package enables a security team or assessor to understand root cause, fix direction, and remediation priority.
Retestability	the degree to which the package and artifacts support later fix verification or repeated testing.
Process Integrity	the truthfulness and integrity of scope adherence, safety controls, assistance disclosure, collaboration disclosure, and evidence handling.

APAS-CORE-4.1.1
APAS SHALL quantify pentest quality primarily through one or more of the quality dimensions defined in this section.

APAS-CORE-4.1.2
Any Q2, Q3, Q4, R2, or R3 claim SHALL include a quality declaration.

APAS-CORE-4.1.3
A quality declaration SHALL represent each applicable quality dimension using a numeric value, ordinal class, boolean status, or other profile-defined representation, and SHALL reference the supporting evidence or methodology.

APAS-CORE-4.1.4
Comparative validation MAY supplement quality assessment but SHALL NOT replace coverage fidelity, finding validity, or evidence traceability.

APAS-CORE-4.1.5
Higher execution autonomy, lower human participation, or public benchmark performance SHALL NOT by themselves satisfy any quality dimension.

APAS-CORE-4.1.6
Where a quality dimension is not applicable or not measured, that absence SHALL be explicitly disclosed.

4.2 Coverage declaration

APAS-CORE-4.2.1
Any T2A, T2C, T3, or T4 claim, and any Q2, Q3, Q4, R2, or R3 claim, SHALL include a coverage declaration.

APAS-CORE-4.2.2
At minimum, a coverage declaration SHALL identify:

Required field	Minimum declaration content
Scope basis	scope source or target inventory reference where material
Reachable surface	reachable attack surface identified during testing
Exercised areas	asset classes, interfaces, roles, tenants, workflows, state transitions, and privilege boundaries exercised where material
Access and context	vantage points, credentials, and context sources used
Coverage status	areas tested, partially tested, excluded, materially untested, and not reachable under the declared conditions
Coverage rationale	basis for material coverage statements where the applicable methodology requires it
Limiting conditions	materially limiting conditions, including timebox, safety restrictions, environmental failures, missing access, or broken prerequisites
Evidence links	references to supporting artifacts or logs

APAS-CORE-4.2.3
A coverage declaration SHALL be truthful about the difference between tested scope, declared scope, and reachable scope.

APAS-CORE-4.2.4
No package SHALL imply complete, full, or equivalent coverage absent declared methodology and supporting evidence sufficient for that representation.

APAS-CORE-4.2.5
Where source code, repository context, implementation artifacts, or comparable context are used, the coverage declaration SHALL distinguish implementation-aware testing from claims of code review completeness unless the applicable profile expressly permits that broader claim.

APAS-CORE-4.2.6
Any Q4 or R3 claim made under BB or GB SHALL disclose why richer context was not used and SHALL state the resulting coverage and evidence limitations.

APAS-CORE-4.2.7
Any T2C claim SHALL structure the coverage declaration so that a relying reviewer can identify control-relevant tested areas, exclusions, and materially untested areas without vendor-side analytical interpretation.

4.3 Finding states

APAS recognizes the following finding states:

Finding state	Meaning in APAS
Observation	An anomalous signal, artifact, or behavior not yet validated as a finding.
Validated Finding	A finding for which the candidate system demonstrates exploitability or strong evidence of a vulnerable condition, pending independent confirmation under the declared methodology.
Confirmed Finding	A finding for which the evaluator or applicable conformance process independently reproduces the vulnerable condition, exploit path, or equivalent supporting evidence sufficient for decision-grade reliance within the declared methodology.
Retested Finding	A prior confirmed finding that has been evaluated again after claimed remediation, with the retest result explicitly disclosed.
Unsupported Finding	A reported issue that is not confirmed under the declared methodology, or whose reported impact is materially overstated after adjudication.

APAS-CORE-4.3.1
Only Confirmed Findings SHALL support Q2, Q3, Q4, R2, or R3 claims.

APAS-CORE-4.3.2
Validated Findings, Observations, and Unsupported Findings MAY be included in a package if clearly labeled, but SHALL NOT be represented as confirmed.

APAS-CORE-4.3.3
A finding whose reported impact is materially overstated SHALL be downrated or treated as unsupported for the overstated portion.

APAS-CORE-4.3.4 Retest results SHALL disclose whether a prior confirmed finding remains present, is partially resolved, or is no longer reproducible under the declared retest conditions. Because model-driven candidate systems are non-deterministic, a retest or reproduction SHALL distinguish reproduction of the vulnerable condition at the target (expected to be stable from a fix) from reproduction of the candidate system's discovery of it (probabilistic). The attestation SHALL disclose materially relevant non-determinism controls and observed run-to-run variability (e.g., number of runs, decoding configuration, variance in findings), and a negative or non-reproducing result SHALL NOT be interpreted as a fix or as absence absent this disclosure.

4.4 Evidence classes

APAS recognizes the following evidence classes:

Evidence class	Meaning
EV1	Signal - Evidence of anomalous behavior, misconfiguration, or suspicious condition.
EV2	Validation - Evidence demonstrating exploitability or a strong vulnerable condition.
EV3	Traceable Confirmation - Evidence that the evaluator or applicable conformance process independently reproduced the vulnerable condition, exploit path, or equivalent proof sufficient for the declared methodology.
EV4	Impact Substantiation - Evidence or adjudicated rationale supporting the represented impact band.

APAS-CORE-4.4.1A Confirmed Finding SHALL include EV3 or equivalent traceable confirmation acceptable under the applicable methodology. Where the methodology permits sampled rather than exhaustive reproduction, the sampling basis SHALL be disclosed, and the evidence package SHALL disclose to the relying party the evaluator engagement model, including who performs and funds independent reproduction and the expected turnaround, so that recurring (e.g., quarterly) reliance can be assessed for cost and timeliness.

APAS-CORE-4.4.2
A High or Critical Confirmed Finding SHOULD include EV4 or evaluator-approved impact rationale.

APAS-CORE-4.4.3
Evidence classes MAY be extended by profiles but SHALL NOT be weakened for Q4 or R3 claims.

4.5 Negative-result discipline

APAS-CORE-4.5.1
A package with no Confirmed Findings, or with only Low severity confirmed findings where higher assurance or reliance is claimed, SHALL include a negative-result statement.

APAS-CORE-4.5.2
A negative-result statement SHALL identify the exercised roles, workflows, privilege boundaries, exclusions, materially untested areas, limiting conditions, and evidence references relevant to interpreting the absence of confirmed findings.

APAS-CORE-4.5.3
No package with no Confirmed Findings SHALL be represented as assessor-ready, complete-coverage, or substitution-ready solely because the report is quiet, short, or contains no material findings.

5. Claim Model

5.1 Claim types

APAS defines three claim types:

Claim type	Meaning
Atomic System Capability Certification	An evaluator-issued result stating that a candidate system satisfied the applicable APAS requirements for one specific capability combination under declared conditions.
Multi-Configuration Capability Statement	A structured statement covering two or more atomic APAS claims for the same candidate system version.
Engagement Conformance Statement	A statement that a specific engagement was executed using an APAS-certified system under the same or narrower declared conditions.

APAS-CORE-5.1.1
An Engagement Conformance Statement SHALL NOT be represented as certification of the target environment itself.

APAS-CORE-5.1.2
A Multi-Configuration Capability Statement SHALL NOT be interpreted as certifying any combination not explicitly listed with its status.

5.2 Claim format

Every APAS atomic claim SHALL use this format:

APAS <Mode>-<Tier> / <Execution>-<Collaboration>-<Outcome> / <Environment Class> [<Affordance Profile>] {<Reliance Outcome>}

Example:

APAS BB-T2C / X2-C2-Q4 / E3 [P-021] {R2}

Meaning:

Claim dimension	Example value	Meaning
Mode	BB	black-box mode
Tier	T2C	compliance evidence pentest tier
Execution class	X2	agent-led execution with in-band human collaboration permitted
Collaboration class	C2	agent-led testing with active human collaboration
Assurance outcome class	Q4	assessor-ready assurance outcome
Environment class	E3	application plus internal adjacency environment class
Affordance profile	P-021	under declared affordance profile P-021
Reliance outcome	R2	intended for augmentation use

APAS-CORE-5.2.1
A public or private APAS claim SHALL include all eight dimensions shown in the required claim format.

APAS-CORE-5.2.2
Each unique claim tuple of mode, tier, execution class, collaboration class, assurance outcome class, environment class, affordance profile, and reliance outcome constitutes a separate capability combination.

APAS-CORE-5.2.3A candidate system version MAY hold multiple atomic APAS certifications, and each SHALL be evaluated independently; certification of one combination SHALL NOT be cited as evidence for any other.

APAS-CORE-5.2.4
Where more than one atomic certification is represented for the same candidate system version, the publisher SHALL provide a capability matrix.

APAS-CORE-5.2.5
Any public or private claim represented as Q2, Q3, Q4, R2, or R3 SHALL include or reference the applicable quality declaration.

5.2A Combination status classes

A capability matrix SHALL use one of the following statuses for every listed combination:

Status	Meaning
S1 Supported	the vendor states the system is designed to operate in this combination, but APAS evaluation has not been completed
S2 Evaluated	the combination was evaluated, but no certification or conformance claim is represented
S3 Certified or Conformant	the combination satisfied the applicable APAS requirements and may be claimed
S4 Not Certified	the combination was evaluated and did not satisfy the applicable APAS requirements
S5 Not Evaluated	the combination has not been evaluated
S6 Unsupported	the vendor does not support this combination for the candidate system version
S7 Retired	the combination was previously supported or certified but is no longer active for the current candidate system version

APAS-CORE-5.2A.1
Any public matrix that omits materially relevant combinations supported by the candidate system version is misleading.

APAS-CORE-5.2A.2
A vendor SHALL NOT represent S1, S2, S4, S5, S6, or S7 combinations as certified.

APAS-CORE-5.2A.3
A public claim SHALL distinguish S3 from all other statuses using clear labeling.

5.2B Capability matrix requirements

A capability matrix SHALL include at minimum, for each listed combination:

Required field	Minimum content
Combination ID	Unique identifier for the listed capability combination
Atomic claim string	full atomic APAS claim string
Candidate system	candidate system name and version
Methodology	methodology reference
Evaluator	evaluator identity where applicable
Status	status class

Required field	Minimum content
Evidence link	attestation ID or evidence reference where applicable
Quality link	quality declaration reference where applicable
Coverage link	coverage declaration reference where applicable
Validity metadata	certification date and validity date where applicable
Limitations	material limitations
Dynamic behavior note	notes on dynamic behavior or mode-selection rules where material

APAS-CORE-5.2B.1
An atomic APAS claim remains the unit of certification. A capability matrix is an indexing and disclosure structure, not a replacement for atomic claims.

APAS-CORE-5.2B.2
If a combination varies by any claim dimension, it SHALL be represented as a separate row or entry.

5.3 Modes of testing

Mode	Description
BB	Black-box testing with no source code and no privileged architecture knowledge beyond declared credentials or materials.
GB	Grey-box testing with limited credentials, documentation, or partial implementation context.
WB	White-box or code-aware testing with source code, repository access, repository-derived context, architecture detail, or deep implementation context.

APAS-CORE-5.3.1
A claim made under WB SHALL NOT be generalized to GB or BB.

APAS-CORE-5.3.2
A claim made under BB or GB SHALL NOT imply equivalent completeness, efficiency, coverage fidelity, signal quality, or evidence quality to a WB claim.

APAS-CORE-5.3.3
Where source code, repository context, implementation detail, or comparable lawful context is available, profiles targeting higher-assurance outcomes SHOULD prefer WB over GB and BB.

APAS-CORE-5.3.4
Any Q4 or R3 claim made under BB or GB SHALL disclose why richer context was not used and SHALL state the resulting limitations.

5.4 Tiers

Tier	Description
T1	Surface assurance.
T2A	Recurring assurance pentest.
T2C	Compliance evidence pentest.
T3	Business-critical deep assessment.
T4	Threat-led resilience profile.

APAS-CORE-5.4.1
The detailed requirements for each tier SHALL be defined in the applicable APAS profile documents.

APAS-CORE-5.4.2T4 SHALL NOT be represented as a generic pass or fail substitution class in the Core standard. A T4 claim SHALL declare the emulated threat-actor profile, the engagement objectives or flags, and whether detection and response were in scope; a T4 claim that excludes social-engineering, physical, and wireless initial-access vectors SHALL disclose that the resilience assessment is limited to the in-scope technical attack surface and SHALL NOT imply full adversary emulation.

APAS-CORE-5.4.3T2C SHALL be interpreted as a pentest evidence tier intended to support compliance review and SHALL NOT by itself imply control pass status, framework acceptance, or assessor acceptance. A T2C claim SHALL reference at least one published framework crosswalk annex identifying the specific framework, control area, and acceptance conditions under which the package is intended to be relied upon; in the absence of any such annex, the claim SHALL NOT use the label 'compliance evidence pentest'.

5.5 Execution autonomy classes

Execution autonomy describes who drives the testing workflow.

Class	Label
X1	Human-led assisted execution.
X2	Agent-led execution with in-band human collaboration permitted.
X3	Independent agent execution.

APAS-CORE-5.5.1: X1 Human-Led Assisted Execution
Humans direct the testing workflow and the candidate system materially assists execution, exploration, or analysis.

APAS-CORE-5.5.2: X2 Agent-Led Execution with In-Band Human Collaboration Permitted
The candidate system drives the testing workflow, but humans MAY provide in-band analytical input, steering, creativity, escalation handling, or report-shaping assistance.

APAS-CORE-5.5.3: X3 Independent Agent Execution The candidate system drives the testing workflow without in-band analytical human assistance. An X3 coverage declaration SHALL state which classes of testing the system attempted and which it is known not to attempt (for example business-logic abuse, multi-step authorization flaws, or creative exploit chaining), so that independent execution is not represented as comprehensive testing. Human activity MAY be limited to evaluator adjudication, administrative clarification, or relying-reviewer consumption after execution; however, a human or automated safety-intervention capability with authority to halt the run SHALL remain available throughout execution. Non-analytical safety intervention does not invalidate an X3 claim.

APAS-CORE-5.5.4
Higher execution autonomy SHALL NOT by itself imply better findings, better evidence, or better pentest quality.

5.6 Collaboration classes

Collaboration classes describe how humans and agents work together.

Class	Label
C1	Human-led with agent assistance.
C2	Agent-led with active human collaboration.
C3	Agent-led with human escalation only.
C4	Agent-only until relying-party review.

APAS-CORE-5.6.1: C1 Human-Led with Agent Assistance
Humans remain the primary testing operators and use the candidate system as an assistant.

APAS-CORE-5.6.2: C2 Agent-Led with Active Human Collaboration
The candidate system leads the workflow, but humans and agents may repeatedly hand off work, reasoning, or next actions during testing.

APAS-CORE-5.6.3: C3 Agent-Led with Human Escalation Only
The candidate system leads the workflow and humans intervene only at defined escalation points, safety events, or evaluator-directed checkpoints.

APAS-CORE-5.6.4: C4 Agent-Only Until Relying-Party Review
The candidate system completes the workflow without human participation other than evaluator oversight or relying-reviewer review after completion.

5.7 Assurance outcome classes

Assurance outcome describes the strength and usability of the resulting pentest output.

Class	Label
Q1	Signal-grade.
Q2	Decision-useful.
Q3	Reference-validated.
Q4	Assessor-ready.

APAS-CORE-5.7.1: Q1 Signal-Grade
The output provides useful technical signal but is not sufficient for decision-grade pentest reliance.

APAS-CORE-5.7.2: Q2 Decision-Useful
The output is independently evaluable, materially reproducible, accompanied by a coverage declaration and limitations statement, and useful for remediation or security decision-making within the declared scope.

APAS-CORE-5.7.3: Q3 Reference-Validated
The output satisfies the applicable evaluator requirements for comparative validation against a declared reference process, reference cohort, or equivalent baseline under the same declared conditions.

APAS-CORE-5.7.4: Q4 Assessor-Ready
The output satisfies the applicable evaluator requirements for a complete assessor-consumable evidence package, including a quality declaration, coverage declaration, traceable finding support, limitations, and negative-result discipline where applicable, without vendor-side analytical interpretation or editing.

APAS-CORE-5.7.5
Q4 SHALL NOT imply framework acceptance by default.

APAS-CORE-5.7.6
Q3 or Q4 MAY be achieved at different execution autonomy or collaboration classes, subject to the applicable profile and evaluator result.

APAS-CORE-5.7.7
Higher assurance outcome classes SHOULD favor the richest lawful context available, including code-aware and implementation-aware testing, where the objective is to reduce noise, improve finding quality, improve coverage fidelity, or accelerate meaningful validation.

APAS-CORE-5.7.8
Public benchmark or public corpus performance SHALL NOT by itself satisfy Q3, Q4, or R3.

APAS-CORE-5.7.9
Q3 is not a prerequisite for Q4 unless the applicable profile expressly requires comparative validation.

APAS-CORE-5.7.10
Comparative validation SHALL NOT authorize omission of a quality declaration, coverage declaration, or evidence traceability requirements.

5.8 Reliance outcomes

Outcome	Label
R1	Evidence-only.
R2	Assessor-consumable augmentation.
R3	Assessor-reliable substitution.

APAS-CORE-5.8.1: R1 Evidence-Only
The output MAY be used as supplemental evidence and SHALL NOT be represented as a substitute for a required pentest activity.

APAS-CORE-5.8.2: R2 Assessor-Consumable Augmentation
The output MAY be used to augment assurance, coverage, retesting, or remediation validation where the package is directly usable by the relying reviewer.

APAS-CORE-5.8.3: R3 Assessor-Reliable Substitution
The output MAY be represented as substituting for a subset of pentest activities only where the applicable profile, evaluator result, quality declaration, and relying framework or assessor permit such use.

APAS-CORE-5.8.4
The Core standard SHALL NOT require a single execution or collaboration class for Q3, Q4, or R3. Any stricter condition SHALL be imposed by the applicable profile.

APAS-CORE-5.8.5
R3 SHALL be interpreted as conditional and profile-bound, not universal.

APAS-CORE-5.8.6
The Core standard does not assume that a stronger result in one claim dimension transfers to any other dimension.

APAS-CORE-5.8.7
No claim under WB transfers to GB or BB. No claim under X3 transfers to X2 or X1. No claim under C4 transfers to C3, C2, or C1. No claim under Q4 transfers to Q3, Q2, or Q1. No claim under R3 transfers to R2 or R1. Any permitted carry-forward SHALL be explicitly defined by the applicable profile or Evaluator Guide.

APAS-CORE-5.8.8 Any R2 or R3 claim SHALL identify the subset of pentest activities, coverage areas, or decision uses to which the reliance claim applies.

6. Core Roles

6.1 Evaluator

APAS-CORE-6.1.1 For X2, X3, Q3, Q4, or R3 claims, the evaluator SHALL be organizationally independent from the vendor. Organizational independence means no common ownership or control, no material revenue dependence on the vendor beyond the evaluation engagement fee, and full disclosure of any financial, equity, or referral relationship between the evaluator and the vendor in the attestation.

6.2 Relying reviewer

APAS-CORE-6.2.1
For any engagement or certification represented as R2 or R3, the intended human reviewer SHALL be the relying auditor, assessor, or customer-appointed reviewer rather than a vendor-side pentest reviewer.

APAS-CORE-6.2.2
Review performed by the relying reviewer after execution SHALL NOT be treated as human assistance.

6.3 Reference validation method

APAS-CORE-6.3.1
Any Q3 claim that relies on human or other comparative validation SHALL be supported by an evaluator-defined reference methodology documented in the Evaluator Guide.

APAS-CORE-6.3.2
Where a human reference cohort is used, the applicable methodology SHALL disclose the cohort qualification rules and conflict controls in the Evaluator Guide.

6.4 Vendor-side analytical review

APAS-CORE-6.4.1
Any vendor-side analytical review used during execution, finding formation, report drafting, quality-declaration formation, coverage-declaration formation, evidence-package assembly, or framework or control mapping SHALL be disclosed.

APAS-CORE-6.4.2
Undisclosed vendor-side analytical review invalidates any Q3, Q4, or R3 claim and any affected quality-declaration fields.

7. Declared Conditions

7.1 Affordance profile

APAS-CORE-7.1.1
Every APAS claim SHALL reference a declared affordance profile.

APAS-CORE-7.1.2
At minimum, the affordance profile SHALL declare:

Required field	Minimum declaration content
Access flags	internet access status; browser access status; terminal or shell access status; source-code or repository access status
Credentials	credential type
Reachability	network reachability boundaries
Vantage points	vantage points tested
Asset coverage	asset classes covered
Tools	tool and plugin inventory where material
Context inputs	context sources provided, including source code, repositories, architecture documents, logs, test credentials, or implementation artifacts where material
Scope basis	target inventory or scope source reference where material
Persistence	memory persistence where material
Safety controls	safety restrictions
Timebox	timebox
Exclusions	exclusions and non-testable areas
Methodology	methodology reference and version

Required field	Minimum declaration content
Mode logic	mode-selection logic where the platform can operate in more than one testing mode
Handoff triggers	escalation triggers and handoff triggers where humans may materially intervene
Dynamic switching	conditions under which the system may switch context sources, credentials, or collaboration pattern during execution
Reconfiguration	whether dynamic reconfiguration is allowed during a run

7.2 Environment classes

Environment class	Description
E1	Single application or API.
E2	Multi-role or multi-step workflow environment.
E3	Application plus internal adjacency, segmentation, or cross-system chaining.

APAS-CORE-7.2.1 All APAS claims SHALL be limited to the declared environment class. Where E3 is claimed, the coverage declaration SHALL state the maximum depth of cross-system chaining actually demonstrated (e.g., single pivot, multi-hop lateral movement, or trust-boundary or segmentation defeat) and SHALL NOT imply deeper chaining than was exercised.

7.3 Methodology declaration

APAS-CORE-7.3.1
Any T2A, T2C, T3, or T4 claim SHALL reference a declared methodology.

APAS-CORE-7.3.2
Any Q2, Q3, Q4, R2, or R3 claim SHALL reference a declared methodology that defines required coverage-declaration rules, required evidence, exclusions handling, and limitations handling.

APAS-CORE-7.3.3
Any methodology targeting Q4 or R3 SHALL define negative-result handling, remediation expectations, and retest expectations.

APAS-CORE-7.3.4
Any methodology targeting Q4 or R3 SHALL state the rationale for the selected mode and SHALL explain whether richer implementation context was available, used, or omitted.

7.4 Dynamic operating behavior

APAS-CORE-7.4.1
Where a candidate system can change mode, execution class, collaboration class, context source, or assistance condition during a run, the evaluator SHALL record the observed operating path.

APAS-CORE-7.4.2 The final attested claim SHALL reflect, for each affected claim dimension, the truthful class that is most conservative for relying-party purposes given what was materially used during the run. For execution autonomy, collaboration, assurance, and reliance dimensions this is the weakest class materially used; for mode, the claim SHALL reflect the richest context materially used and SHALL NOT be treated as a quality downgrade relative to lower-context modes.

APAS-CORE-7.4.3
If source code, repository-derived context, or comparable implementation context is materially used during a run, the claim SHALL be attested under WB for that run.

APAS-CORE-7.4.4
If in-band analytical human assistance materially changes strategy, exploitation, validation, prioritization, or report content during a run, the run SHALL NOT be attested as X3.

APAS-CORE-7.4.5
If the work pattern includes repeated human-agent analytical handoff, the run SHALL NOT be attested as C3 or C4.

APAS-CORE-7.4.6
A candidate system MAY receive separate atomic claims for separately bounded operating configurations, provided those configurations are independently declared and evaluated.

APAS-CORE-7.4.7
The coverage declaration and quality declaration SHALL reflect any materially relevant dynamic changes in context, assistance, or mode.

8. Evidence Package Requirements

8.1 Minimum package contents

APAS-CORE-8.1.1
Any APAS certification or engagement statement represented as R2 or R3 SHALL include an assessor-consumable evidence package.

APAS-CORE-8.1.2
Any Q2, Q3, or Q4 claim SHALL include at minimum a quality declaration and coverage declaration, even where R1 is claimed.

APAS-CORE-8.1.3
The minimum evidence package SHALL contain:

Required element	Minimum content
Declaration	certification declaration or engagement declaration
Reliance statement	reliance outcome statement
Quality declaration	quality declaration
Methodology	methodology statement
Coverage declaration	coverage declaration
Mode and context statement	mode-selection and context-usage statement

Required element	Minimum content
Scope summary	scope and rules-of-engagement summary
Context metadata	environment class and affordance profile reference
Combination reference	combination ID where applicable
Matrix reference	capability matrix reference where the candidate system has more than one listed combination
Finding-state summary	finding-state summary, including confirmed findings and unsupported findings where applicable
Confirmed findings	confirmed findings summary
Evidence reference	evidence manifest or evidence bundle reference
Exclusions	exclusions and untested areas statement
Limitations	limitations statement
Dynamic path summary	observed operating path summary where dynamic behavior is material
Truthful attestation note	explanation of why the attested combination truthfully reflects the run where dynamic behavior is material
Remediation reference	remediation guidance reference or roadmap where findings exist
Negative-result statement	negative-result statement where applicable
Retest information	retest information where applicable
Attestation file	machine-readable attestation
Compliance mapping reference	control or requirement mapping reference where required by the applicable T2C profile or crosswalk annex

APAS-CORE-8.1.4
For a multi-configuration candidate system, the evidence package SHALL permit the relying reviewer to determine exactly which combination was evaluated, which combinations are certified, and which combinations remain unevaluated or unsupported.

APAS-CORE-8.1.5
Any T2C package represented as Q4, R2, or R3 SHALL permit the relying reviewer to determine what control-relevant areas were tested, excluded, or materially untested without vendor-side analytical interpretation.

8.2 Package integrity

APAS-CORE-8.2.1
The evaluator or applicable conformance process SHALL preserve sufficient traceability to link reported findings and coverage declarations to supporting artifacts.

APAS-CORE-8.2.2 Fabricated, non-traceable, or materially misleading evidence invalidates the claim. Because LLM-driven candidate systems are prone to hallucinated findings, fabricated exploit chains, and invented or misattributed CVE or reference identifiers, EV3 independent reproduction SHALL be performed against the target condition itself and SHALL NOT consist solely of replaying the candidate system's own asserted narrative or steps. Every cited external identifier (e.g., CVE, advisory) SHALL be independently verified to exist and to match the reported condition.

APAS-CORE-8.2.3
A materially misleading coverage declaration or negative-result statement invalidates the affected quality-declaration fields and any affected reliance claim.

8.3 Assessor consumability

APAS-CORE-8.3.1
A package represented as R2 or R3 SHALL be structured so that the relying reviewer can evaluate scope, coverage, findings, limitations, and the declared reliance outcome without vendor-side analytical interpretation.

APAS-CORE-8.3.2
For any Q4 claim, the package SHALL be complete without vendor-side analytical editing.

APAS-CORE-8.3.3
Where a claim is made under BB or GB, the package SHALL disclose the likely limitations arising from reduced context and SHALL NOT imply equivalent code awareness.

8.4 Machine-readable attestation

APAS-CORE-8.4.1 Any Q2, Q3, Q4, R2, or R3 package SHALL include machine-readable fields for the following. The machine-readable attestation SHALL be cryptographically signed by the evaluator (and, where the vendor produces fields, separately signed by the vendor), and the evidence manifest SHALL list a content hash for each referenced artifact so that any post-hoc alteration of the attestation, logs, or evidence is detectable. Run logs and the observed operating path SHALL be captured in a tamper-evident manner (e.g., append-only or hash-chained) by a component not under the sole control of the candidate system. The required fields are:

Field group	Required machine-readable field
Claim identity	full atomic claim string
Combination identity	combination ID where applicable
Candidate system	candidate system name and version
Methodology	methodology reference
Evaluator	evaluator identity where applicable
Assistance disclosure	assistance class

Field group	Required machine-readable field
Dynamic behavior	observed operating path summary or reference where dynamic behavior is material
Quality link	quality declaration reference
Coverage link	coverage declaration reference
Finding summary	finding-state summary
Evidence link	evidence-manifest reference
Validity metadata	dates and validity fields where applicable
Reliance metadata	reliance outcome

9. Human Assistance and Collaboration Integrity

9.1 Assistance classes

Assistance class	Meaning
H0	No human assistance affecting strategy, exploitation, validation, prioritization, or reporting.
H1	Administrative or safety-only intervention.
H2	Analytical assistance. H2 SHALL be sub-classified as H2a strategic/exploitation assistance (vulnerability hypothesis, exploit-chain design, lateral-movement or pivot decisions, attack-path selection) and H2b interpretive/editorial assistance (prioritization, control mapping, report editing). H2a SHALL be disclosed distinctly from H2b.

APAS-CORE-9.1.1 Any claimed assistance class SHALL be disclosed in the attestation. For any X2, X3, C3, or C4 claim, the attestation SHALL additionally disclose the underlying model or model class powering autonomous decision-making, the safeguards applied to it, and known failure modes material to safe autonomous operation against live targets, including susceptibility to prompt injection, jailbreaking, or social engineering by the target environment.

APAS-CORE-9.1.2 Any X3 claim SHALL prohibit H2 (analytical) assistance during execution. Non-analytical safety intervention (H1), including invocation of the stop/suspend mechanism, SHALL remain permitted and available throughout an X3 run and does not invalidate the X3 claim.

APAS-CORE-9.1.3
Any Q4 claim SHALL prohibit H2 during evidence-package assembly.

APAS-CORE-9.1.4
Use of H2 does not by itself invalidate X1, X2, C1, C2, Q1, or Q2 claims, provided it is accurately disclosed.

APAS-CORE-9.1.5
For multi-configuration systems, assistance disclosure SHALL be combination-specific and, where relevant, run-specific.

APAS-CORE-9.1.6
Any post-execution vendor-authored framework mapping, prioritization, or conclusion-shaping that materially affects the relying reviewer SHALL be treated as vendor-side analytical review and disclosed.

9.2 Collaboration truthfulness

APAS-CORE-9.2.1
The declared collaboration class SHALL reflect how work was actually performed.

APAS-CORE-9.2.2
Undisclosed handoffs, undisclosed analytical intervention, or materially misleading collaboration claims invalidate the affected claim dimensions.

APAS-CORE-9.2.3
A platform SHALL NOT be marketed as agent-only, independently autonomous, or assessor-ready where materially relevant runs depended on undisclosed human switching, human review, vendor-side control mapping, or post-execution analytical reshaping for the represented combination.

10. Core Safety and Integrity Requirements

APAS-CORE-10.1 An APAS claim SHALL require explicit scope boundaries and an effective stop/suspend mechanism. The mechanism SHALL: (a) be invocable by the target owner and the evaluator, not only the vendor; (b) halt the candidate system within a declared bounded reaction time; (c) interrupt in-flight actions, including partially executed exploits, to the extent technically feasible; (d) fail safe by halting offensive activity if the agent loses contact with its oversight channel; and (e) be tested and verified by the evaluator as a precondition of the claim. The affordance profile SHALL declare the stop authority, reaction-time bound, and loss-of-oversight behavior.

APAS-CORE-10.2
An APAS claim SHALL require preservation of sufficient logs or artifacts for evaluator review.

APAS-CORE-10.3 Out-of-scope activity, fabricated evidence, material concealment of assistance, or materially misleading coverage statements invalidate the claim. In addition, the candidate system SHALL enforce technical scope-boundary controls (e.g., network egress allow-listing, target allow-listing, and action gating) designed to PREVENT out-of-scope action during execution. Invalidation of a claim is a disclosure consequence and SHALL NOT be treated as a substitute for a preventive out-of-scope containment control. Any out-of-scope action that occurs SHALL be reported to the target owner and evaluator as a safety incident regardless of claim status.

APAS-CORE-10.4
The detailed safety tests, replay requirements, and failure conditions SHALL be defined in the Evaluator Guide or the applicable profile.

APAS-CORE-10.5 Material uncertainty in impact, coverage, or retest conclusions SHALL be explicitly labeled in the evidence package. Safety controls SHALL scale with execution autonomy and collaboration independence: any X3 or C4 claim SHALL additionally require documented pre-deployment adversarial safety evaluation of the candidate system (including loss-of-control, scope-escape, and prompt-injection/jailbreak resistance against hostile targets); continuous independent monitoring of the run with an enforced stop authority; and a defined incident-response and target-owner notification procedure. Higher autonomy SHALL NOT be evidenced solely by stronger output assurance.

11. Public Claims and Marketing Language

11.1 Required disclosure

APAS-CORE-11.1.1
Any public APAS claim SHALL disclose:

Disclosure item	Required public disclosure
Claim form	whether the claim is atomic or matrix-based
Claim dimensions	mode; tier; execution class; collaboration class; assurance outcome class; environment class; affordance profile; reliance outcome
Evaluator	evaluator identity
Dates	certification date; validity date
Candidate version	candidate system version
Limitations	material limitations where reliance is claimed
Quality availability	availability of the applicable quality declaration or a summary thereof
Adjacent combinations	the status of adjacent combinations that a reasonable buyer may infer are available, where the candidate system supports more than one combination
Dynamic switching	whether the candidate system supports dynamic switching among combinations, where material

11.2 Prohibited claims

APAS-CORE-11.2.1
No party SHALL state or imply any of the following solely on the basis of APAS certification:

Prohibited implication
universal human equivalence

Prohibited implication
universal pentest substitution
framework-wide compliance
assessor acceptance by default
substitution without declared limitations and represented subset
higher pentest quality solely because execution autonomy is higher
equivalence between BB or GB testing and WB testing unless that equivalence is explicitly justified and declared
platform-wide capability where only one combination is certified
implied certification of adjacent or related combinations that have not been certified
a misleadingly incomplete capability matrix that hides unsupported, retired, failed, or unevaluated combinations
complete or full coverage absent supporting coverage declaration and methodology
high assurance, assessor readiness, or substitution readiness solely because a report has few or no findings
benchmark leadership or public-corpus performance as proof of Q3, Q4, or R3

APAS-CORE-11.2.2
No party SHALL market an execution class, collaboration class, assurance outcome class, or reliance outcome without the associated full claim dimensions.

12. Governance and Companion Documents

12.1 Companion documents

The APAS document set SHOULD include at minimum:

Companion document	Purpose
APAS Core Standard	Core claim semantics and minimum normative concepts.
APAS Evaluator Guide	Evaluation procedures, scoring, safety testing, and adjudication rules.
APAS Profile Registry	Objective-specific and framework-aligned profile requirements.
APAS Attestation Schema	Machine-readable fields for claims, quality, coverage, and evidence.

Companion document	Purpose
optional framework-specific annexes	Framework crosswalks, mapping notes, and additional reliance conditions where applicable.

APAS-CORE-12.1.1
Where this Core document defers implementation detail, the applicable companion document SHALL define the binding requirement.

APAS-CORE-12.1.2
Where this Core document requires quality declarations or coverage declarations, the APAS Attestation Schema SHALL define the required machine-readable fields.

12.2 Versioning

APAS-CORE-12.2.1
Major version changes SHALL be used for changes to claim semantics, quality-dimension semantics, conformance structure, execution classes, collaboration classes, assurance outcome classes, or reliance outcomes.

APAS-CORE-12.2.2
Minor version changes SHOULD be used for new definitions, non-breaking requirements, or additional profile support.

APAS-CORE-12.2.3 Patch versions SHOULD be limited to errata and clarifications. A material change to the candidate system's underlying model, agent architecture, or capability profile SHALL be treated as a new candidate system version and SHALL trigger re-evaluation of any affected atomic claim before that claim may continue to be represented as certified. The Core standard and Evaluator Guide SHALL be reviewed for capability drift on a defined cadence not exceeding twelve months, and SHALL define a trigger for out-of-cycle review when frontier autonomous-pentest capability materially shifts.

12.3 Public review

APAS-CORE-12.3.1
Normative changes to the Core document SHOULD undergo public review before release.

APAS-CORE-12.3.2APAS SHALL be maintained through an open, vendor-neutral, versioned governance process under a neutral steward independent of any single vendor, with published conflict-of-interest rules for the maintaining body and for evaluator accreditation, public comment tracking, and public disposition of substantive comments. The author of any candidate system SHALL NOT accredit, employ, or control the evaluator of its own claims.

13. Interpretation Rules

APAS-CORE-13.1
If a conflict exists between this Core document and a profile, the profile MAY add stricter requirements but SHALL NOT weaken Core claim semantics or Core quality-dimension semantics.

APAS-CORE-13.2
If a conflict exists between a public claim and the attested declaration, the attested declaration governs.

APAS-CORE-13.3
If a framework or assessor requires conditions stricter than APAS, those stricter conditions govern for that use.

APAS-CORE-13.4
If execution class, collaboration class, assurance outcome, quality declaration, and reliance outcome appear to conflict, the evaluator result and attested declaration SHALL determine which dimension governs the claim.

APAS-CORE-13.5
If a public marketing summary conflicts with a quality declaration or coverage declaration, the attested quality declaration and coverage declaration govern.

Appendix A: Example Claim Declarations

Example claim	Reading of the declaration
APAS BB-T2A / X1-C1-Q2 / E1 [P-014] {R1}	Human-led pentest with agent assistance, decision-useful output with declared coverage and limitations.
APAS WB-T3 / X2-C2-Q3 / E2 [P-031] {R2}	Agent-led code-aware testing with active human collaboration, reference-validated output.
APAS WB-T2C / X2-C2-Q4 / E3 [P-021] {R2}	Agent-led testing with repeated human-agent handoff, rich implementation context, assessor-ready output, augmentation use.
APAS BB-T2C / X3-C4-Q4 / E3 [P-022] {R3}	Independent agent execution, no in-band human collaboration, assessor-ready output, conditional substitution use. This claim SHALL require explicit disclosure of why richer context was not used, the resulting limitations, and the represented subset of pentest activities.

Appendix B: Companion Documents

The following documents are expected to accompany this Core draft:

Companion document	Scope
Evaluator Guide	scoring, run procedures, safety tests, comparative validation methods, combination-level evaluation methods, coverage-fidelity evaluation methods, and anti-gaming controls
Profile Registry	T1, T2A, T2C, T3, T4, and framework-specific profiles, including mode-selection guidance, context-rich testing requirements, and any permitted carry-forward rules where applicable

Companion document	Scope
Attestation Schema	machine-readable declaration fields, combination identifiers, quality-declaration fields, coverage-declaration fields, finding-state fields, observed operating path fields, capability-matrix status fields, and evidence-manifest requirements
Crosswalk Annexes	framework-specific reliance notes, mapping rules, and limitations

Appendix C: One-Sentence Core Standard

An autonomous system may claim APAS-certified pentest capability only when, within a declared mode, tier, execution class, collaboration class, assurance outcome class, environment class, affordance profile, and reliance outcome, it produces bounded or continuously-scoped pentest evidence whose coverage fidelity, finding validity, evidence traceability, remediation usability, retestability, and process integrity can be evaluated by a relying reviewer within the limits stated by the applicable APAS profile and attestation, with explicit disclosure of the context used, the materially tested and untested scope, and the specific capability combination actually evaluated.

Appendix D: Informative declaration headings

A quality declaration for Q2, Q3, Q4, R2, or R3 will typically include at minimum:

Quality declaration component	Typical content
Quality-dimension summary	quality-dimension summary for coverage fidelity, finding validity, evidence traceability, remediation usability, retestability, and process integrity
Coverage link	coverage-declaration reference
Finding-state counts	finding-state counts
Unsupported finding disclosure	unsupported-finding disclosures where applicable
Evidence link	evidence-manifest reference
Remediation and retest	remediation and retest references where applicable
Integrity disclosures	assistance and vendor-side analytical review disclosures
Limitations	limitations and uncertainty summary

A coverage declaration will typically include at minimum:

Coverage declaration component	Typical content
Scope basis	scope source or inventory reference

Coverage declaration component	Typical content
Reachable surface	reachable attack-surface summary
Exercised paths	roles, workflows, and privilege boundaries exercised
Coverage status	tested, partially tested, excluded, materially untested, and unreachable areas
Access and context	credentials, vantage points, and context sources used
Limiting conditions	materially limiting conditions
Evidence links	supporting artifact references
Negative-result link	negative-result reference where applicable