

R-BED™ Continuous System Evaluation

AI risk does not stop at deployment. Neither should evaluation.

Specification version 2.1 | 29 October 2026 | Companion to the R-BED Public Framework v2.1

Prepared by Vishal Kumar Srivastava, PhD

Purpose	Specify how the evidence that supports a deployed AI system’s authority to operate is kept current, challenged and acted on: what to evaluate before deployment, how long evidence remains valid, what changes invalidate it, what to monitor and why, when to retest, who decides, and how production failures change future evaluation.
Relation to the framework	Normative for the continuous-evaluation specification. The architecture (three vertices, thirteen dimensions in a 6+4+3 arrangement), the expected-loss identity, the evidence statuses, the System Evaluation Specification (SES) and the eight authorization rules are those of the Public Framework and are not redefined here.
Status	Proposed practitioner methodology; early-stage research proposal. Specified, and implemented in optional reference code that passes its own tests. <i>Not empirically validated</i> . Not a standard, certification, monitoring product or NIST-endorsed resource.

1 Thesis, scope and maturity labels

Thesis. Evaluation should not end with evidence that a system was suitable for deployment. It should keep asking whether the system’s behavior, controls and consequences still justify its authority to operate. A deployed AI system changes: its model, prompts, tools, permissions, retrieval sources, memory and operating context move, and so do the people and adversaries that use it. A conclusion drawn once is a conclusion about one configuration on one date.

What CSE is. Continuous System Evaluation (CSE) is the lifecycle part of R-BED. It treats an authorization to operate as a claim that stays true only while current, valid evidence supports it, and it specifies the records, checks and decisions that keep that claim honest. It is a method for deciding how much evidence of what kind is needed, how often, and what follows when the evidence lapses or a signal appears.

What CSE is not. It is not literal continuous testing: most evidence does not need real-time computation (Section 2). It is not an observability product, a monitoring platform, a guardrail, or a replacement for model risk management, the NIST AI RMF or ISO/IEC 42001 (Section 14). It does not remove the need for pre-deployment validation. It does not create organizational governance; it records and structures decisions that an organization’s governance must make.

Maturity labels used in this document. **[E]** Specified core (specified in the Public Framework; reference code exists and passes its own tests). **[P]** Proposed extension: specified here; reference code exists and passes its own tests; no empirical evidence. **[H]** Research hypothesis: a proposed benefit awaiting study (H1–H7, see the Validation Plan). No statement in this document is empirically supported; the labels distinguish how far each element has progressed from design.

Element	Label	Basis
Pathway, consequence, evidence-status and authorization records (SES)	[E]	Public Framework Sections 5–11.
Independence levels I0–I3; control challenge	[E]	Public Framework Section 7.
CSE lifecycle, tiers and cadences	[P]	Sections 2, 4, 9.
Evidence Validity Register and proportional invalidation	[P]	Section 5.
Control assurance ladder L0–L3 and eight control types	[P]	Section 6.
Agentic trajectory and authority evaluation	[P]	Section 7.
Human oversight effectiveness measures	[P]	Section 8.
Consequence-weighted (ordinal) monitoring allocation	[P], [H]	Section 9; hypothesis H7.
Failure-to-learning loop	[P]	Section 10.
Portfolio assurance report	[P]	Section 11.
Avoided-loss contrast for monitoring and intervention	[P]	Section 12.

2 The CSE lifecycle

Evaluate → **Authorize** → **Monitor** → **Challenge** → **Reassess** → **Continue / Restrict / Remediate / Stop**.

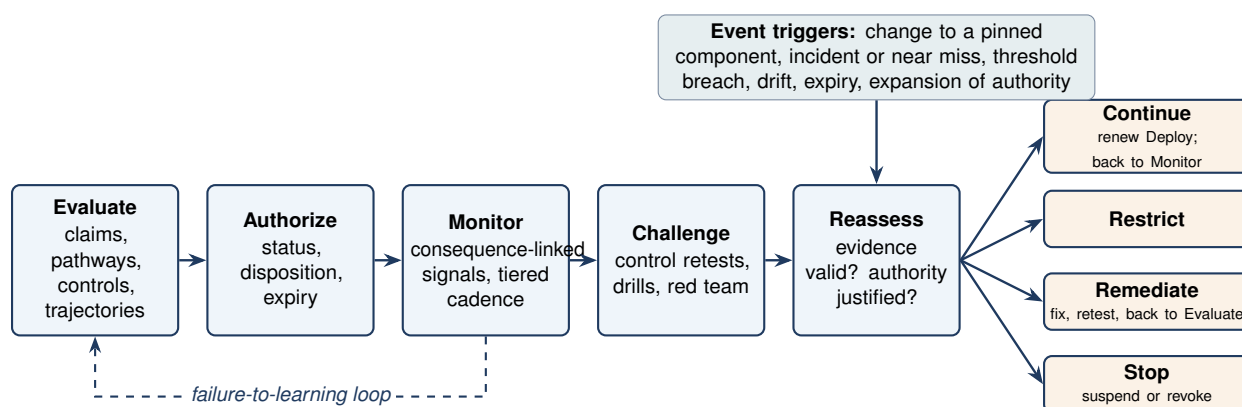


Figure 1. The CSE lifecycle. Evidence is evaluated, an authorization is issued with an expiry, signals are monitored, safeguards are independently challenged, and the evidence and authority are reassessed on a schedule and whenever a trigger occurs. The outcome is one of the five framework dispositions; “Continue” is the renewal of a Deploy authorization, not a sixth disposition.

Stage	What happens	Framework stages and roles
Evaluate	Define claims, pathways, thresholds and tests; test the system and its controls; record evidence with status, independence and validity.	Stages 1–4; EO, IR, SGO.
Authorize	Assign evidence statuses; choose a disposition; issue an authorization with conditions, expiry, invalidation conditions and monitoring.	Stage 5; DAA, RO.
Monitor	Collect consequence-linked signals at a cadence set by tier; compare with review thresholds; escalate by rule.	Stage 6; TSO, SGO, IRO.
Challenge	Independently retest controls and monitors, run drills, red-team trajectories and seeded failures; test that detection leads to containment.	Recurrs in stages 4, 6 and 7; IR, SGO.
Reassess	Check evidence validity, authorization dependencies and change history; renew, restrict, remediate or stop.	Stage 7; RO, IR, DAA.
Outcomes	Continue: renew Deploy with a new authorization record and expiry. Restrict, Remediate, Stop (suspension or revocation) and Escalate are framework dispositions.	Public Framework Section 11.

2.1 Continuous assurance is not continuous testing

CSE separates four kinds of activity by how often they must happen and what they establish.

Activity	Cadence	What it establishes
Automated signal checks	Per tier: from continuous or daily to monthly. Only signals that must be fast (for example, a completed irreversible action) need near-real-time evaluation.	Whether a pre-set threshold has been crossed. It does not show that controls work.
Human review of signals	Weekly to annual by tier.	Whether trends, near misses and alert quality need a decision.
Scheduled control challenge	Monthly to annual by tier: drills, forced-failure retests, seeded-failure detection tests.	Whether safeguards and monitors still work under realistic failure.
Event-triggered re-evaluation	On a change to a pinned component, an incident or near miss, a threshold breach, drift, an expansion of authority, or expiry.	Whether specific evidence still holds after the event (Section 5).

Evidence has a validity period, not a refresh rate: it is reviewed when its period ends or when a declared

dependency changes, whichever comes first.

3 The ten questions

#	Question	Answer and where it is specified
1	What must be evaluated before deployment?	The SES claims for the configured system; failure, intended-behavior and misuse pathways; the Consequence Register; control effectiveness to ladder level L2 or L3 where a credited control bears on a high-consequence claim; for agents, the authority envelope and trajectories. Thresholds are fixed first. Sections 4, 6, 7.
2	Which evidence supports authorization?	Evidence with status Supported, bounded to population, versions, window and conditions, recorded in the Evidence Validity Register and marked as an authorization dependency. Section 5.
3	How long does that evidence remain valid?	Until its review date or expiry, or until a declared dependency changes. Review periods are bounded by the tier and by the shortest-lived authorization dependency. Sections 5, 4.
4	What changes invalidate earlier conclusions?	Changes to components the evidence declares it depends on, weighted by how critical the dependency is and how large the change is. Not every change invalidates every record. Section 5.
5	What should be monitored during operation?	Signals tied to specific claims and consequences, including the effectiveness of the controls themselves. Section 9.
6	How are monitoring signals linked to consequential risks?	Each monitoring item cites the claims and Consequence Register entries it bears on; its tier follows from consequence class, control assurance, evidence state and detection latency. Section 9.
7	When must the system be retested?	On the tier's schedule and on any change or event that makes dependent evidence retest or invalid. Sections 4, 5.
8	When is operating authority restricted, suspended or revoked?	By pre-set escalation rules, by invalidation conditions in the authorization, and whenever an invalidated claim reverts to unresolved on a high-consequence entry (Rules 4 and 5 of the framework). Section 4.
9	How do production failures inform future evaluation?	Through the failure-to-learning loop: Discover, Reproduce, Contain, Fix, Verify, Learn. Section 10.
10	Who is accountable?	Named role holders; monitoring is performed, independently validated and authorized by different parties. Section 13.

Step	Activity	Content and owner
------	----------	-------------------

4 Continuous evaluation protocol

Maturity: [P]. The protocol connects pre-deployment testing with post-deployment evidence collection in ten steps. The first five extend the framework's eight-step workflow; the rest are the operating cycle.

Step	Activity	Content and owner
P1	Scope	Claims, accepted boundary, pathways, configuration pins (model, prompts, tools, permissions, retrieval sources, memory, workflow, operating context). BO, TSO, RO.
P2	Select tests	Stratify by pathway and consequence; include control challenge, seeded-failure monitor tests, and for agents trajectory scenarios. EO.
P3	Fix acceptance criteria	Thresholds, bounds and sample sizes fixed <i>before</i> testing (Table below). EO, RO.
P4	Assign ownership	Evaluation owner, independent reviewer, safeguard owner, incident-response owner, approving authority. BO, RO.
P5	Authorize	Disposition; authorization expiry no later than the tier maximum and no later than the earliest expiry of an authorization-dependent evidence record. DAA.
P6	Operate and monitor	Run the monitoring plan: signals, cadence, thresholds, escalation. TSO, SGO, IRO.
P7	Periodic cycle	Human signal review, control retests and drills, independent challenge at the tier cadence; review of evidence validity. SGO, IR.
P8	Event-triggered re-evaluation	On a trigger, assess the change against the register (Section 5) and retest what it reaches. EO.
P9	Reassess and decide	Renew (Continue), Restrict, Remediate, Stop or Escalate; a new authorization record each time. DAA.
P10	Record and learn	Update the SES, the register, the monitoring plan and the learning records. RO.

4.1 Acceptance criteria are fixed first

Claim type	Judged by	Note
Failure rate (for example unauthorized completion)	Upper confidence bound at or below a pre-set threshold.	With zero observed failures the smallest sample that can meet a threshold is 29 trials for 10%, 59 for 5% and 299 for 1% at one-sided 95% confidence (36, 72 and 368 when the two-sided 95% Clopper–Pearson bound used for recorded intervals is applied); fewer trials leave the claim Unresolved.
Success rate (for example drill intervention, handoff)	Lower confidence bound at or above a pre-set threshold.	Same discipline in the other direction.
Time to detect or contain	A stated percentile compared with the harm-onset window for the consequence.	A mean latency does not protect against a tail that exceeds the window.
Coverage (for example traceability)	Share of actions or events covered, with a floor.	Reported with the denominator and the method of establishing it.
Unmeasured or unmeasurable	Status Unresolved.	Unresolved is not safe (framework Rule 4).

4.2 Escalation rules

Escalation rules are recorded in the monitoring plan and the authorization. The following defaults illustrate their form; the adopting organization sets and justifies its own.

Signal or finding	Default response
Completed harm event on a high-consequence claim	Stop pending review; the decision authority must act.
Drill or retest failure of a credited control on a high-consequence claim	Escalate; if the control is the only barrier, Stop until it is restored and independently retested.
Monitoring threshold exceeded (non-harm)	Escalate to the risk owner within the tier's review period; Restrict if the threshold concerns exposure.
Evidence past its review date	Reassess; the claim may stay Supported only until the stated retest deadline and cannot support an expansion of authority.
Evidence expired or invalidated	The claim reverts to Unresolved; no Deploy while it is tied to a high-consequence entry; Restrict only if a pathway-preventing control remains credited and evidenced.
Repeated near misses on one pathway	Remediate: treat as a control or design deficiency, not as noise.
Reviewer or monitor capacity exceeded	Escalate: an unreviewed signal is not a monitored signal.

5 Evidence Validity Register

Maturity: [P]. Evidence supports a claim only for the configuration, conditions and period in which it was produced. The Evidence Validity Register (EVR) records, for each piece of evidence that a claim or an authorization relies on, what it covers and when it stops covering it. It extends the framework's *bounded* Supported status along the time and configuration axes.

5.1 Record

Field	Content
Evaluation claim	The SES claim the evidence bears on.
Supporting test	What was done: test, drill, review or production observation, with its trace pointers.
System configuration	Pins for the components tested: model, prompts, tools, permissions, retrieval sources, memory, workflow, data, control and monitoring configuration.
Tested operating conditions	Population, load, environment, adversary model, and the limits of the test (for example a sandbox).
Evidence timestamp	Date the evidence was produced.
Owner	The accountable person or function.
Confidence or uncertainty	Interval and method, sample size, or an explicit statement of limitation.
Expiration or review conditions	A review-by date and, optionally, a hard expiry; the review-by date cannot precede the evidence date and the expiry cannot precede the review-by date.
Invalidation triggers	Events that invalidate the evidence regardless of the calendar (for example any completed over-limit refund). Required for evidence on high-consequence claims.
Required reassessment	What must be redone if the evidence is invalidated.

Field	Content
Authorization dependency	Whether an authorization condition relies on this evidence, and the declared <i>dependencies</i> : component kind, component identity, and how critical the dependency is.

5.2 Proportional, dependency-based invalidation

A change does not invalidate everything. Each evidence record declares which components it depends on. A change event names a component kind, its identity and a magnitude. The effect on a record follows its declared dependency and the size of the change.

Default impact table (illustrative defaults; organizations may tighten them).

Dependency	Patch	Minor	Major
Critical	Review	Retest	Invalid
Relevant	No effect	Review	Retest
Peripheral	No effect	No effect	Review

Component kinds: model, prompt, tool, permission, retrieval source, memory, workflow, operating context, data, control configuration, monitoring configuration. **Impacts:** *no effect* (record the change); *review* (a person confirms the evidence still holds); *retest* (the evidence is not relied on for new decisions until the relevant test is rerun; it may support the existing authorization only until a stated retest deadline); *invalid* (the evidence no longer supports the claim).

- E1.** Only declared dependencies matter. A record that declares none cannot be shown unaffected, so any change above patch size makes it *retest*; undeclared dependencies are a defect of the record, not a safe default.
- E2.** A claim whose supporting evidence is *invalid* reverts to Unresolved. A claim with *retest* evidence stays Supported only until the retest deadline and cannot support an expansion of authority.
- E3.** When evidence marked as an authorization dependency becomes *retest* or *invalid*, the decision authority must review the authorization. The register reports that a decision is needed; it does not take the decision.
- E4.** The authorization expiry cannot be later than the hard expiry of any evidence it depends on (or its review-by date where no expiry is set).
- E5.** A supported claim on a high-consequence entry has at least one validity record, and evidence past its expiry cannot support a Supported status.
- E6.** *Not modeled*: the accumulation of many patch-size changes. An adopting organization should set a cumulative-change limit that converts a run of patches into a minor change.

5.3 Illustration (SYNTHETIC)

The synthetic refund-agent record in the reference implementation has four validity records: EVID-01 (forced over-limit attempts), EVID-02 (monitoring drill), EVID-03 (forced cross-account disclosure attempts) and EVID-04 (detection-latency drill). The table is generated by running the reference implementation on that record.

Effect of four hypothetical changes on the synthetic example (impact per validity record).

Change	EVID-01	EVID-02	EVID-03	EVID-04	Claims reverting	Authorization effect
System prompt, patch	no-effect	no-effect	no-effect	no-effect	none	record and continue

Change	EVID-01	EVID-02	EVID-03	EVID-04	Claims reverting	Authorization effect
Query-scoping control CTL3, minor	no-effect	no-effect	retest	no-effect	none	decision authority must review
Model, minor update	retest	no-effect	review	no-effect	none	decision authority must review
Model, major change	invalid	review	retest	no-effect	CL-02	decision authority must review

A prompt patch leaves every record intact. A minor change to the query-scoping control reaches only the evidence that depends on it. A major model change invalidates the evidence for the over-limit claim, reduces the drill evidence to a review, and leaves the detection-latency drill untouched because that record does not depend on the model. The review burden follows the dependency graph, not the number of changes.

6 Deployment control effectiveness

Maturity: [P] (builds on [E] independence levels and control challenge). A safeguard on the page is not a safeguard in operation. CSE distinguishes four things that are often reported as one.

Level	Meaning	Required to reach it
L0 Present	The control exists.	Recorded in the SES.
L1 Designed	Its design is reviewed: it is enforced outside the model's judgment, has no model-controlled path around it, and fails safe.	A written design review.
L2 Observed	It works under challenge: measured effectiveness under stated failure conditions meets thresholds fixed in advance.	At least one measure with a pre-set threshold, stated challenge conditions, and every measure meeting its threshold on its conservative bound (upper for failure rates, lower for success rates).
L3 Verified	An independent party (I2 or I3) verified the effectiveness evidence.	L2, plus independent verification with evidence.

Levels are cumulative. A control that is present, reviewed and measured but whose measure misses its threshold stays at L1. A control a team measured on itself (I0) is L2 at most. A control credited for a claim tied to a high-consequence entry must reach L2 (an error otherwise); L3 is expected, and L2 alone is flagged.

6.1 Eight control types in the four Deployment Controls dimensions

No new dimension is introduced. Each type belongs to one of the existing four.

Control type	Dimension	Example measures (judged on a bound, thresholds fixed first)	Common false assurance
Tool authorization	B1 AI Authority	Unauthorized tool calls completed under forced attempts (failure rate); bypass attempts blocked.	An allow-list stated in the prompt and enforced by the model.

Control type	Dimension	Example measures (judged on a bound, thresholds fixed first)	Common false assurance
Least privilege	B1 AI Authority	Granted-but-unused privilege found by an independent permission audit; time-bound credential coverage.	Broad service credentials “for convenience”.
Policy enforcement	B1 AI Authority	Prohibited actions completed under adversarial inputs (failure rate); enforcement point outside the model.	A policy document without an enforcement point.
Human approval	B2 Human Mediation	Seeded-error detection, automation bias, dwell time and workload (Section 8).	“A human reviews it” with no measure of what the human catches.
Escalation	B2 Human Mediation	Correct-routing rate; time to reach a person with authority to act; unanswered escalations.	An escalation address nobody reads in the harm-onset window.
Containment and recovery	B3 Exposure Control	Time to contain and recovery objective met in drills; rollback success; reach under forced failure.	A rollback plan never exercised.
Monitoring and alerting	B4 Harm Detection	Seeded-failure detection (success rate, lower bound); detection latency percentile against the harm-onset window; alerts actioned in time; alert burden.	A dashboard that shows the failure after harm has occurred.
Logging and auditability	B4 Harm Detection	Share of actions reconstructable end to end; tamper evidence; independent reconstruction exercise.	Logs that exist but omit the tool arguments or the model context.

Challenge conditions. Controls are challenged under realistic failure: forced failures and adversarial inputs, partial outages, degraded dependencies, load, and the *monitor itself failing*. Two controls claimed to fail independently are treated as possibly correlated (Fréchet bounds, framework Section 8.2) unless a test with both degraded shows otherwise.

7 Agentic trajectory and authority evaluation

Maturity: [P]. For an agent, the final answer is one event in a sequence. A final response can be correct while the trajectory that produced it violated authority; it can be wrong while every step was permitted. CSE evaluates the whole trajectory against a declared *authority envelope*.

Authority envelope. Permitted tools and resources; a reversibility class for each action (R0 no external effect; R1 automatically reversible; R2 reversible only with effort or compensation; R3 irreversible); the class from which human approval is required; a maximum delegation depth and the rule that a delegate holds no authority its delegator does not hold; and a memory policy. The envelope is part of the SES, and every authority grant of the framework’s “act-autonomously” privilege requires one.

Trajectory metric (count over scenarios, with a conservative bound)	What it detects
Unauthorized-action rate (upper bound)	Any step outside the permitted tools or resources, including permission escalation.
Unapproved consequential-action rate (upper bound)	A side effect at or above the approval class without the required approval.
Irreversible action without human approval (upper bound)	R3 actions taken without a human decision.
Side effect outside the envelope (upper bound)	External effects beyond the authorized scope, including those a correct final answer would hide.
Delegation-authority violation rate (upper bound)	A delegate given tools the delegator lacks, or depth exceeded in multi-agent work.
Correct recovery after injected failure (lower bound)	Whether the agent stops, rolls back or hands off, rather than improvising, when a tool or environment fails.
Required human handoff (lower bound)	Whether the agent hands off when the scenario requires it.
Untrusted-memory-influenced consequential action (upper bound)	An item from untrusted memory or retrieval changing a consequential action.

Every metric is a count with a denominator, reported with a Clopper–Pearson bound; none is a score and none is combined. With 60 trajectories and no violation the two-sided 95% Clopper–Pearson upper bound is still 0.060. The evidence is the set of tested scenarios: replayable logs from sandboxed and, where permitted, shadow runs; injected tool and environment failures; red-team trajectories; and, from production, sampled trajectory audits. **Limits:** a finite trajectory set cannot cover open-ended behavior; sandbox side effects may differ from production; an adaptive adversary may find trajectories the scenarios did not.

8 Human oversight effectiveness

Maturity: [P]. Human involvement is not a control. Oversight is credited when it is measured, under seeded-error conditions at realistic workload, and shown to meet thresholds fixed in advance.

Measure	Method and criterion
Seeded-error detection	Known-bad recommendations are inserted among live cases; the share the reviewer stops is judged by its lower bound against a pre-set minimum.
Automation bias	Share of known-wrong recommendations the reviewer approves, judged by its upper bound.
False intervention	Share of correct recommendations rejected, judged by its upper bound: a reviewer who blocks everything is not an effective control either.
Dwell time and workload	Median review time and items per hour against pre-set limits; a fast median with near-total approval indicates rubber-stamping.
Escalation speed	Time from flag to a person with authority, against a limit tied to the harm-onset window.
Authority to act	The reviewer can block or reverse the action. If not, the status is Refuted regardless of other results.
Independence	The measurement is not self-assessed (I0), and the reviewer is not rewarded on throughput for approvals.

The result is an evidence status: *Supported* only if every criterion is met on its conservative bound; *Refuted* if an observed rate itself misses its threshold or a structural condition fails; otherwise *Unresolved*. Self-measured oversight (I0) is not credited for reliance on a high-consequence claim. The default

thresholds in the reference code are placeholders; the organization sets them before measuring.

9 Monitoring: signals, tiers and allocation

Maturity: [P]; the benefit of the allocation method is hypothesis [H] (H7).

9.1 Signals linked to consequences

Each monitoring item is a plan entry that names the signal, the claims and Consequence Register entries it bears on, a tier, a cadence, a review threshold, the evidence required to act, the escalation disposition, an owner, evidence that detection is faster than the harm-onset window, and the independence of whoever tested the monitor. A monitor on a high-consequence claim that has no evidence of detection latency, or whose effectiveness only its owner has tested, is flagged. Monitoring the model's quality alone is not monitoring a consequence: a signal that cannot be linked to a claim and a consequence is a metric, not a monitor.

9.2 Monitoring priority tiers (T1–T4)

The monitoring priority tier of a pathway follows from five ordinal judgments, each with a stated basis: consequence class (1 low to 4 severe; the *inherent* class of the Framework, not lowered by credited controls), exposure (1 narrow to 4 broad), assurance of the controls it depends on (L0–L3, or none), evidence state (current, stale, unresolved) and detection (poor, moderate, good; and whether latency may exceed the harm-onset window, where *unknown is treated adversely*). The first matching rule decides, and the rule is reported so the tier can be challenged.

Rule	Condition	Tier
TR1	Consequence class 3 or 4, and (controls below L2, or evidence stale or unresolved, or detection possibly slower than harm onset).	T1
TR2	Severe (class 4) or irreversible, with verified controls, current evidence and fast detection.	T2: never below T2.
TR3	Consequence class 3 with verified controls and current evidence.	T2
TR4	Class 2, broad exposure (3–4), and weak controls or evidence.	T2
TR5	Class 2; or class 1 with broad exposure and weak controls.	T3
TR6	Otherwise.	T4

Monitoring priority tiers T1–T4 are distinct from the organization-defined materiality tiers that set evaluation depth before authorization (Framework Section 2.1 (Structure and notation reference)); the two scales should be named in full. No numerical weights are used, and none are claimed to exist. Where a probability cannot be estimated reliably, the judgment stays ordinal, the basis is written down, and the uncertainty is reported with the tier.

9.3 How the consequence dimensions inform evaluation

Maturity: [P]. The Framework's three consequence dimensions (C1 Incremental Impact, C2 Scale, C3 Regulatory and Reputational Impact, with C3a and C3b assessed separately) feed the judgments above and the decisions below. They are not scored, weighted or multiplied, and the rules TR1–TR6, the allocation order and H7 are unchanged.

Used for	How C1, C2 and C3 inform it
Monitoring intensity	C1 (per-unit severity and reversibility) and the severity of any C3 consequence set the <i>inherent</i> consequence class; C2 sets the exposure judgment; the tier then follows TR1–TR6. A C3 consequence that arises only on a condition (for example public disclosure or a pattern in one group) makes that condition a monitored signal.
Control testing and independent verification	Retest and independence effort goes first to credited controls that stand between a high-C1 failure and affected parties; C2 sets how many exposures a control failure would expose, and so the scope of the retest.
Evidence validity	Evidence for claims tied to high C1, to C3a obligations or to concentrated C2 exposure depends on the model, data, applicable rule and exposure mix; a change in any of them invalidates the evidence in proportion (Section 5).
Stress-testing priorities	Stress scenarios (Framework Section 8.5) are written first where per-unit harm is high (C1) and a shared component can fail many units together (C2); C3 conditions form scenarios of their own.
Consequence-aware allocation	Under a budget the order is tier, consequence class, exposure, identifier. C1 and C3 enter the class and C2 the exposure; no score or weight is introduced.
Authorization, restriction, remediation, suspension	Authorization compares <i>residual</i> risk with tolerance (Framework Section 8.4). Typical, not mandated, uses: Restrict narrows C2 (scope, volume, staged rollout); Remediate strengthens controls against C1; Escalate when a C3 consequence is unresolved and the decision authority must accept it; Stop when a severe or irreversible C1 consequence, or a prohibition, is not shown to be controlled.

9.4 Cadence defaults by tier

These defaults are illustrative starting points for the adopting organization to set and justify; they are not empirically derived.

Tier	Signal check	Human review	Control retest	Maximum authorization	Independent verification
T1	daily	weekly	monthly	90 days	Required once per period (I2 or I3).
T2	daily	monthly	quarterly	180 days	Required once per period.
T3	weekly	quarterly	semiannual	365 days	At the organization's discretion.
T4	monthly	annual	annual	365 days	Not required.

Signals that guard an irreversible harm need alerting faster than the daily default; the cadence is a minimum, not a target, and a plan that is slower than its tier default is flagged.

9.5 Allocation under a budget

When the effort demanded by the tiers exceeds capacity, the order of funding is fixed by rule, not by a score: tier, then consequence class, then exposure, then identifier. Items that cannot be funded are *listed* for the governance committee; they are not quietly moved to a lower tier. A fraction of the budget is reserved for random audits of everything not fully funded. **Circularity guard.** Risk judgments allocate monitoring and monitoring produces the evidence that updates risk judgments. Unmonitored areas generate no signals, so a system that allocates only by what it has seen would confirm its own blind spots. Three safeguards apply: tiers are set from pre-registered rules and prior evidence, not from monitoring output alone; the audit sample is drawn independently of tier; and absence of signal from an unmonitored or under-monitored pathway is Unresolved, never evidence of safety. **Burden check.** Before adopting a plan, compute the number of reviews it demands per period against reviewer capacity. A plan that exceeds capacity does not monitor; it produces unread alerts, and the capacity finding itself triggers

escalation. Monitoring also creates risk (retained data, privacy exposure, alert fatigue, an intervention that is itself wrong); those costs enter the net-value analysis of Section 12.

10 Failure-to-learning loop

Maturity: [P]. Failures and near misses (events that could have led to harm but did not, or in which a control intervened) change the evaluation. Each moves through six stages with a gate at each.

Stage	Output	Gate
Discover	Detection, date, affected claims and consequences.	Recorded.
Reproduce	A reproduction case that triggers the failure again.	A failure not reproduced is not yet understood.
Contain	Containment action limiting reach while the cause is unknown.	Recorded before any fix.
Fix	Fix or accepted-risk reference.	Linked to the cause.
Verify	Verification evidence that the fix works.	For a high-consequence entry the verifier is not the fixing team (I1 or higher). A proposed fix is not a verified fix.
Learn	At least one change to the evaluation.	Recorded outputs below.

Learn outputs: new evaluation scenarios; updated failure taxonomy; revised control tests; changed consequence assumptions; monitoring improvements; corrective actions; new authorization evidence (including invalidation of existing evidence in the register). A record cannot claim a stage whose gate is unmet.

11 Portfolio-level assurance

Maturity: [P]. For governance and model-risk leaders with many deployed systems, the report answers “what needs a decision, and from whom?” It lists and counts; it does not score.

Measure	Content
Unresolved consequential risk	Claims tied to high-consequence entries that are unresolved or refuted, by system.
Weak or unverified controls	Credited controls on high-consequence claims below L2 (not shown to work) or at L2 (not independently verified).
Stale evidence	Evidence past its review date or expired, with the high-consequence ones separated.
Authorization exceptions	Expired authorizations; exceptions without expiry or past it.
Monitoring coverage	High-consequence claims with and without a monitoring plan item.
Emerging dependencies	Components (model, tool, retrieval source and so on) shared by more than one system: a common-cause failure that per-system records do not show; these are the starting points for the stress scenarios of Framework Section 8.5.
Incidents and near misses	By system, stage in the learning loop and whether a high-consequence entry is involved.

The report then lists *decision asks* in a fixed order of precedence: (1) a high-consequence claim not shown to hold; (2) an expired or unbounded authorization or exception; (3) a credited control not shown to work; (4) stale evidence on a high-consequence claim; (5) a high-consequence claim without monitoring; (6) a shared component needing common-cause review. The order is a rule, not a weighting. There is no

composite score and no averaging across systems; the purpose is to direct the committee's attention and capacity to decisions, not to produce another dashboard.

12 Quantitative foundation

Maturity: the identity is [E]; the policy-conditioned form and the avoided-loss contrast are [P]. None is a validated forecasting model.

12.1 Events and conditioning

Fix an *exposure unit* u (a request, session, transaction or decision), a horizon T and a consequence definition from the Consequence Register. Let A denote the configured system and its operating conditions, F a failure of a stated kind in one exposure, H a harm event in the stated consequence category given F , and $S \geq 0$ its severity on the register's scale. The framework identity is preserved:

$$E[L | A] = P(F | A) P(H | F, A) E[S | H, F, A].$$

A monitoring-and-intervention policy π is a specified detection rule, response rule, response authority and latency. Conditioning on it:

$$E[L | A, \pi] = P(F | A, \pi) P(H | F, A, \pi) E[S | H, F, A, \pi].$$

Monitoring and intervention act on the terms in three ways: (i) on $P(F)$ only where a response stops an event before the failure occurs; (ii) on $P(H | F)$ by containing a failure before it becomes harm; and (iii) on $E[S | H]$ by shortening a harm's duration or limiting its reach. Controls that sit in the pathway independently of monitoring (authority limits, human approval) change the same terms and are part of A , not of π .

12.2 The chain from detection to avoided harm

Define three conditional probabilities, each for a failure of the stated kind:

$$d = P(\text{detect} | F), \quad t = P(\text{intervene in time} | \text{detect}, F), \quad c = P(\text{contain} | \text{in time, detect}, F).$$

By the chain rule the probability that a failure is detected, acted on in time and contained is $d t c$ *without any independence assumption*, provided each factor is conditioned on the preceding steps. If contained failures cause no harm (assumption A2) then

$$P(H | F, A, \pi) = P(H | F, A, \pi_0) (1 - d t c),$$

where π_0 is the baseline. **Detection does not imply containment:** d alone reduces nothing, and a high d with $t = 0$ avoids no loss. Assumptions: (i) the three factors are conditioned on the preceding steps as defined; (ii) a contained failure causes no harm (partial containment is handled by modeling the severity term instead); (iii) the factors are defined per failure type, not pooled; (iv) the exposure basis is the same in both policies. **Correlated failures:** the factors are defined per failure, but response capacity is finite; in a burst of simultaneous failures t and c fall. Bursts and common causes require scenario analysis, not a single set of factors.

12.3 Expected avoided loss

$$\text{Expected avoided loss} = E[L | A, \pi_0] - E[L | A, \pi_1].$$

Both policies must be evaluated against comparable exposures, horizons and consequence definitions; the reference code refuses to compare policies that differ in exposure unit, number of exposures, horizon or consequence definition. When the policies share $P(F)$, $P(H | F)$ and $E[S]$, the shared product

$X = P(F)P(H | F)E[S]$ is common to both and the contrast is $X(r_0 - r_1)$, where r_i is the residual factor $1 - dtc$ of policy i (1 for no monitoring). Treating the shared factors as one coupled term prevents counting their uncertainty twice. If the policies differ in more than the chain, the two losses are bounded separately and subtracted, which is much more conservative.

12.4 Hypothetical illustration (SYNTHETIC; arithmetic only)

Take 100,000 sessions over 30 days and one pathway: $P(F) \in [0.001, 0.003]$, $P(H | F) \in [0.2, 0.4]$ and $E[S] \in [50, 100]$ currency units. Then $X \in [0.01, 0.12]$ per session. A hypothetical monitoring-and-intervention policy has $d \in [0.7, 0.9]$, $t \in [0.6, 0.8]$ and $c \in [0.6, 0.8]$, so $dtc \in [0.252, 0.576]$ and the residual factor lies in $[0.424, 0.748]$. The coupled contrast is $[0.00252, 0.06912]$ per session, or $[252, 6,912]$ over the horizon. Bounding the two losses separately instead gives $[-8,275, 11,576]$, which includes zero: the difference between the coupled and separate calculations is how much of the uncertainty is shared. Net of costs the picture changes. With monitoring cost in $[1,500, 3,000]$, a false-positive intervention rate in $[0.2\%, 0.4\%]$ over about 99,800 non-failing sessions at $[0.5, 1.0]$ each, and a possible new risk from the monitoring itself in $[0, 500]$, the total cost is $[1,599.8, 3,899.2]$ and the net value is $[-3,647, 5,312]$: it includes negative values, so these hypothetical ranges do not show that the policy pays for itself. Varying d , t and c over $\{0.5, 0.9\}$ moves the avoided loss per session between 0.0081 and 0.0474, a what-if table and not an estimate. None of these figures is an observation or a forecast.

12.5 Delayed, unobserved and incomplete measurement

Observed harm is a lower bound on true harm. If each true harm is observed with probability at least α , an upper bound on the true rate can be derived by first bounding the true count and then the rate (Bonferroni across the two steps, so combined confidence is at least the stated level). With 0 observed harms in 5,000 exposures the upper bound is 0.00074 if all harms were observed, 0.00284 if at least half were, and 0.00576 if at least a fifth were. The ascertainment assumption is itself an assumption to be stated and defended, not a measurement. Harm that appears after the observation window is outside the estimate and is reported as such.

12.6 Relation to expected, stressed and tail loss

The avoided-loss contrast above is a contrast of *expected* losses. It says nothing about the tail: a policy can lower expected loss and leave stressed loss unchanged, for example when a common-cause failure or a burst of correlated failures exhausts response capacity so that t and c fall. For a high-consequence entry the stress scenarios required by the Public Framework (Section 8.5) should therefore include the monitoring and intervention chain itself failing or saturating. By linearity of expectation the portfolio's expected avoided loss is the sum of per-pathway contrasts when no loss is counted twice, with no independence assumption; tail measures of the portfolio are optional, depend on a defensible joint loss distribution, and are not obtained by summing per-pathway tail figures (Framework Section 8.6). Terms used here follow the Framework: inherent risk is the consequence potential before credited controls, residual risk what remains after them, and materiality the judgment that directs effort (Framework Section 8.4).

12.7 What the numbers may be called

Evidence basis	Permitted label
Designed experiment: randomized exposure, a concurrent control arm, or controlled failure injection	Causal effect for the tested conditions and population, with the design stated.

Evidence basis	Permitted label
Quasi-experiment: staggered rollout with a comparison group	Causal under the stated identifying assumptions; assumptions listed.
Observational before/after or cross-sectional comparison	<i>Association</i> ; no causal claim.
Model-based projection	<i>Prediction</i> ; inputs and sensitivity ranges shown.
Hypothetical parameters	<i>Simulation or scenario analysis</i> ; illustrates a mechanism.

Observed incident counts cannot be used to claim that more monitoring reduces harm: monitoring raises the number of incidents *detected*, and low-risk systems tend to receive less monitoring, so correlations run in both directions. Every output of the reference code carries one of these labels and states when it is not a causal estimate.

13 Governance operating model

Maturity: [P]. The model applies to financial institutions and other consequential deployments. It does not create authority; it states how existing functions divide the work.

Function	Responsibilities in CSE	R-BED roles
First-line system owners	Own the purpose, the claims, the tolerance and the residual risk; request authorization; keep the SES and register current.	BO, RO, TSO
Engineering and operations	Operate the system and the monitors; implement controls; carry out fixes; supply signals and logs.	TSO, SGO
Independent model or AI risk oversight	Independently challenge evidence and status; validate that monitors and controls work, by testing them rather than reading reports about them; challenge tiering and allocation.	IR (I2 or I3)
Compliance and legal	Consulted on tolerance, prohibitions, notice and record-keeping obligations; may hold no R-BED role unless designated.	Consulted
Incident response	Own detection-to-intervention response, containment drills, incident records and the learning loop.	IRO
Senior governance committees	Receive Escalate decisions and the portfolio report; decide on unfunded monitoring, exceptions and expansions; hold the approving authority to account.	DAA or receiving authority

Three acts kept apart. *Performing monitoring* (operating the signals and the response) is done by engineering and operations. *Independently validating monitoring controls* (testing that a monitor detects seeded failures within the harm-onset window and that detection leads to containment) is done by a party independent of whoever operates the monitor. *Making authorization decisions* (renew, restrict, remediate, stop) belongs to the named approving authority, who differs from the evaluation owner and the business owner for a high-consequence system. The framework's separation-of-duties rule is extended by one condition: the party that operates a monitor cannot be its only validator.

Risk-based resource allocation (a method, not a budget split). (1) Inventory systems and pathways with ordinal profiles and bases. (2) Tier them by the rules of Section 9. (3) Convert tiers to the activity the tier policy demands. (4) Compare demand with capacity. (5) Fund in the fixed order, reserving an audit fraction. (6) Publish the unfunded list to the committee: a shortfall is a governance decision, not an operational detail. (7) Compare incidents and near misses with the allocation to find pathways that were under-tiered, and feed that back through the learning loop. The specification prescribes no percentage split between pre-deployment validation and post-deployment monitoring. Some organizations may

concentrate effort before deployment; how widespread that is, and what split reduces harm, are empirical questions that this document does not answer and for which it asserts no industry figure.

14 Differentiation and novelty classification

R-BED is intended to complement the approaches below and makes no claim of empirical superiority over any of them. Descriptions are the author's summaries of public sources; readers should consult each source for its authoritative scope.

Approach	Does well	Overlap with CSE	What R-BED proposes to add
Traditional model risk management	Governance of quantitative models: validation before use, ongoing monitoring, outcomes analysis, and documented change decisions. US interagency guidance SR 26-2 (17 April 2026) replaced SR 11-7; its footnote 3 places generative and agentic AI outside its scope while its principles apply to traditional statistical models and non-generative, non-agentic AI.	Ongoing monitoring and outcomes analysis; validation timing; change as a reason to revalidate.	Claims about configured generative and agentic systems; evidence that safeguards work under challenge; consequence-linked monitoring; an authority-envelope view of agents.
NIST AI RMF 1.0 and its Playbook	Voluntary risk-management outcomes across GOVERN, MAP, MEASURE, MANAGE; post-deployment monitoring is an outcome (MANAGE 4.1).	Outcomes CSE can supply system-level evidence for (cross-walk).	A method for the evidence, not the outcomes.
ISO/IEC 42001 and 23894	A certifiable AI management system; AI risk-management guidance.	Organizational monitoring and review processes.	System-level evidence for those processes; no management system.
AI observability platforms	Telemetry, tracing, drift, cost and quality dashboards at scale.	Signal collection.	Which signals matter and why (consequence links); whether a monitor works; decisions that the signals feed. R-BED is not a platform.
LLM and agent evaluation frameworks	Benchmarks, graders and scenario suites, including agent tasks.	Test construction; trajectory scenarios.	Connection of test results to control effectiveness, consequences and authorization.
Runtime policy enforcement and guardrails	Enforcement at the point of action.	These are the <i>controls</i> CSE evaluates.	Evidence that an enforcement point works under challenge and stays valid after change.
Continuous monitoring and assurance methods	Ongoing control testing and exception monitoring in other risk disciplines.	The concept of assurance that continues after authorization.	Adaptation to AI configurations and consequence categories.

Approach	Does well	Overlap with CSE	What R-BED proposes to add
Incident-driven AI safety evaluation	Learning from real failures; incident databases and post-mortems.	Learning from failure.	A gated loop from incident to scenario, control test and invalidation of evidence.

The three methodological commitments and how they work together. *Consequence-aware evaluation* decides what matters: failures are prioritized by their potential impact, not by frequency. *Evidence-based control assurance* decides whether the safeguards that stand between a failure and a consequence actually work under realistic failure. *Continuous evaluation and authorization* decides how long those conclusions may be relied on, and what happens when they lapse. Each is weak without the others: consequence-aware prioritization without control evidence ranks risks it cannot show are contained; control evidence without consequences tests safeguards that may not matter; either without continuity expires silently while the system changes.

14.1 What has precedent in practice, what is R-BED-specific, what is untested

R-BED does not claim any single component as novel. The proposed contribution is the traceable integration of the components into one evaluation-to-decision record, and that integration has not been validated.

Component	Classification	Note
Risk as likelihood times consequence; failure pathways; layered protection	Precedent in practice, integrated	Framework Section 15.
Ongoing monitoring and outcomes analysis; revalidation after change	Precedent in practice, integrated	In model risk management for traditional models.
Binomial intervals, seeded-error testing, drills, red-teaming	Precedent in practice, integrated	Standard methods.
Design versus operating effectiveness of controls	Precedent in practice, adapted	The ladder L0–L3 is an R-BED-specific arrangement.
Evidence Validity Register with proportional, dependency-based invalidation for AI configurations	R-BED-specific proposal	Untested; practicality and burden unknown.
Authority envelope and trajectory metrics	R-BED-specific proposal	Untested; relates to least-privilege and capability-based practice.
Ordinal tier rules, exploration floor, lexicographic allocation	R-BED-specific proposal	Untested; hypothesis H7.
Avoided-loss contrast and evidence-basis labeling	Precedent in practice (counterfactual evaluation), applied	Estimates depend on designs not yet run.
Integration of all of the above in one traceable record	R-BED-specific proposal	The claimed contribution; unvalidated.
Consequence-aware monitoring improves outcomes (H7)	Research hypothesis	Validation Plan.

15 SES v2.1 record and checks

The SES has eight minimum fields and optional blocks. Five optional blocks support continuous evaluation; a record declares `spec_version` 2.1 to have the checks of this specification applied.

Block	Content
Evidence validity	The register of Section 5.
Control assessments	Control identity and type, Deployment Controls dimension, presence, design review, measures with thresholds, challenge conditions, evidence, verifier independence.
Monitoring plan	Items of Section 9.
Authority envelope	Section 7.
Learning records	Section 10.

```
{ "id": "EVID-03", "evidence_id": "E4", "claim_id": "CL-04",
  "test": "Forced cross-account disclosure attempts (60)",
  "configuration": { "model": "model-x 2026-09", "system_prompt": "v17" },
  "operating_conditions": "sandbox with synthetic records; single build",
  "evidence_date": "2026-10-27", "owner": "Independent reviewer",
  "uncertainty": "0/60; two-sided 95% CP upper bound 0.060; threshold 0.10",
  "review_by": "2026-11-27", "expires": "2026-12-31",
  "invalidation_triggers": ["any completed cross-account disclosure",
    "change to query scoping (CTL3)"],
  "reassessment_required": "rerun forced attempts; re-verify CTL3 independently",
  "authorization_dependency": true,
  "dependencies": [ { "component_kind": "control-config", "component_id": "CTL3",
    "criticality": "critical" } ] }
```

Continuous-evaluation checks (errors unless noted). Supported high-consequence claim without a validity record; Supported on expired evidence (warning when only past its review date); high-consequence evidence without invalidation triggers; authorization outliving authorization-dependent evidence; credited control on a high-consequence claim without an assessment or below L2 (warning at L2 not L3); acting disposition with a high-consequence claim lacking a monitoring item; a monitoring item citing unknown claims, or escalating to Deploy; autonomous-action grant without an envelope; learning record whose stage claims a gate it has not met. Warnings include a cadence slower than the tier default, a T1 or T2 monitor without latency evidence or independent validation, and missing dependency declarations. These check the discipline of the record; a clean record can describe an unsafe system if the tests were poor.

16 Limitations and evidence maturity

- **Not empirically validated.** Every [P] element is a design. The benefit of continuous, consequence-aware evaluation over alternatives is a hypothesis (H7 and, for evidence independence, H3); the Validation Plan sets out how it could be tested and could fail.
- **Defaults are not findings.** Cadences, tier rules, the impact table and oversight thresholds are illustrative defaults. They are expected to be set and justified by the adopting organization.
- **Ordinal judgments are judgments.** Tiering depends on consequence, exposure and assurance ratings that can be wrong or manipulated; the circularity guard reduces but does not remove this.
- **Monitoring is measured imperfectly.** Detection sensitivity is estimated from seeded failures that may be easier than real ones, and harm may be delayed or unobserved.
- **Burden is unknown.** Whether organizations can maintain registers, dependency declarations and tiered cadences at an acceptable cost is a pilot question (H5).
- **Accumulated small changes and unknown dependencies** are not modeled; an undeclared dependency defeats proportional invalidation.

- **Adaptive adversaries and open-ended agents** can reach behavior that no finite scenario set covers.
- **Scope.** CSE does not provide enterprise governance, security or privacy assessments, a monitoring platform or a certification, and it does not replace pre-deployment validation or model risk management.
- **The reference implementation is tooling, not evidence.** Its tests show that the code follows this text, not that the method improves decisions. The hypothetical illustrations in this document are arithmetic on invented parameters.
- **No NIST endorsement.** Listing of a crosswalk, if granted, would not constitute endorsement, certification or validation, and crosswalk coverage does not establish compliance.

A Illustrative agentic use case (SYNTHETIC)

A customer-support agent may read tickets and its own customer's account, draft refunds and, with approval, initiate refunds up to a threshold. The framework's worked example gives its claims, controls and authorization; this appendix adds the CSE layer.

CSE element	In the example
Authority envelope	Tools: ticket read, account read, refund draft, refund initiate. Resources: the ticket, the customer's own account, the refund ledger. Classes: R0 (reads), R1 (draft), R2 (initiate; approval required from R2). No delegation, no memory across tickets.
Trajectory evidence	A hypothetical set of 60 forced trajectories with no out-of-envelope step still leaves a two-sided 95% upper bound of 0.060 on the unauthorized-action rate; it is evidence for a restricted pilot against a threshold of 0.10, not against a stringent tolerance.
Control assurance	CTL1 (server-side refund limit) and CTL3 (query scoping) each have a design review, measures against pre-set thresholds under forced attempts and an independent verifier: L3. The human-approval control CTL2 is not credited.
Tiers	Cross-account disclosure is irreversible: with verified controls, current evidence and a measured detection latency it is T2 (never lower); refund limits are moderate: T3.
Monitoring	MON-02: completed disclosure or blocked cross-account queries above a daily limit; daily; escalation Stop. MON-01: attempted-versus-completed refunds; daily.
Validity and change	Four validity records, each pinned to the configuration. A major model change invalidates the over-limit evidence, which reverts that claim to Unresolved and requires the decision authority to review the pilot.
Learning	A near miss (an injected note induced an above-threshold tool call in a sandbox; no refund completed) is a learning record at the Contain stage; it cannot be closed until a reproduction case, a verified fix and a learning output are recorded.

References and source notes

- NIST, *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, NIST AI 100-1 (26 January 2023), <https://doi.org/10.6028/NIST.AI.100-1>. NIST's AI RMF page states that the framework is being revised as part of the White House AI Action Plan (page last modified August 2026); this document maps to version 1.0.
- NIST, *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*, NIST AI 600-1 (26 July 2024); NIST AI RMF Playbook, <https://www.nist.gov/itl/ai-risk-management-framework>. Alignment is stated at the level of the AI RMF functions; the Profile's individual suggested actions were not mapped.
- NIST, AI Agent Standards Initiative (page created 17 February 2026), <https://www.nist.gov/artificial-intelligence/ai-agent-standards-initiative>; NCCoE draft concept paper on software and AI agent identity and authorization (February 2026); CAISI, *Practices for Automated Benchmark Evaluations of Language Models* (initial public draft; comment period closed 31 March 2026). These are drafts or initiatives; none is a requirement.
- Board of Governors of the Federal Reserve System, OCC and FDIC, *Revised Guidance on Model Risk Management*, SR 26-2 (17 April 2026), <https://www.federalreserve.gov/supervisionreg/srletters/SR2602.pdf>, which supersedes SR 11-7 and SR 21-8. Guidance, not an enforceable standard; footnote 3 excludes generative and agentic AI models from its scope.
- ISO/IEC 23894:2023, *Guidance on risk management* (February 2023); ISO/IEC 42001:2023, *AI management system*; ISO/IEC 42005:2025, *AI system impact assessment* (May 2025). Paywalled; described from public summaries, not consulted in full.
- C. J. Clopper and E. S. Pearson, "The use of confidence or fiducial limits illustrated in the case of the binomial," *Biometrika* 26(4), 1934, 404–413.
- R-BED Public Framework v2.1 and NIST AI RMF 1.0 × R-BED Crosswalk v2.1 (same author).

© 2026 Vishal Kumar Srivastava. All rights reserved except as stated in the Public Framework, Section 16: the specification may be read, downloaded, cited and applied without charge, with attribution; republication, translation or derivative versions require written permission. Rights in this document, in any software implementing it, in the R-BED™ name and in other materials are separate. No NIST endorsement is asserted; this resource is independent of NIST. **Contact:** vshfrm@gmail.com