The Justification Review Standard (JRS)

A Record-Level Documentation Governance Control for Decision Defensibility in AI-Assisted Environments

Author: Phillip Wikes, Creator of the Justification Review Standard; former Lead Civil Rights Officer, Maryland Commission on Civil Rights

Status: Working research paper, validation phase. Preliminary and observational. No proven-effectiveness claims.

Version: 1.0 · 2026 · jrsstandard.com

Abstract

Organizations increasingly reach consequential decisions with the assistance of generative artificial intelligence, and the documentation that records those decisions increasingly becomes the organization's permanent evidence. A gap has opened between the moment a decision is made and the moment it is later examined, in litigation, audit, investigation, regulatory review, or public-records disclosure. When that later examination occurs, reviewers rarely have access to the people, conversations, or context that produced the decision. They have the record. The Justification Review Standard (JRS) is a record-level documentation governance control that evaluates whether an administrative record can be reconstructed and defended from its own contents, independent of the author's memory and independent of the technology that helped produce it. This paper presents the origins of JRS in civil-rights investigative practice, its conceptual foundation in the distinction between documentation *quality* and documentation *defensibility*, its five review conditions, its treatment of AI-specific documentation risk, its measurement instrumentation, its three-rung validation program and honest current status, and its enterprise implementation architecture. JRS is deliberately positioned not as an AI governance framework, a compliance program, or a legal-sufficiency standard, but as a complementary control that operates on the documentation layer those systems produce.

1. Introduction and Origins

1.1 The problem: Decision Reconstruction Risk

Every consequential decision is made in a context. The people involved understand why it was reached; the evidence is fresh; the reasoning is present in the room. That context is perishable. Months or years later, when the decision is questioned, the surrounding circumstances are no longer visible. Personnel have changed, memories have faded, informal explanations have disappeared. What remains is the written record.

JRS names the failure condition this creates: **Decision Reconstruction Risk (DRR)**, the condition in which a record cannot explain, on its own terms, why a consequential decision was made. A record exhibiting DRR may reflect a sound decision, competently reached, and still fail when examined, because the basis for the decision lived in the author's knowledge rather than on the page. The decision was defensible; the record is not.

This distinction, between the quality of a decision and the defensibility of the record documenting it, is the conceptual center of JRS.

1.2 Origins in civil-rights investigative practice

JRS did not originate in an academic or software context. It emerged from investigative and documentation-review work in the civil-rights enforcement setting, where the adequacy of a written record routinely determines whether an administrative action survives challenge. Investigators, fact-finders, and equal-opportunity counselors repeatedly

confront the same pattern: a determination that reads as complete and professional but whose stated conclusions cannot be traced to identifiable evidence within the file. In that environment, an unanchored characterization is not a stylistic flaw. It is an exposure.

The standard was developed with a cognitive-behavioral and AI-governance lens, drawing on how reviewers actually read records under scrutiny and on the emerging reality that a growing share of records are now drafted with machine assistance. The enforcement origin matters because it establishes the criterion by which JRS judges a record: not whether the record is persuasive, but whether an independent reviewer, arriving with no prior knowledge, could reconstruct the path from evidence to conclusion.

1.3 The shift that makes JRS timely: AI-assisted documentation

Generative AI has changed how records are created. A model produces a fluent first draft, a person edits lightly, and the document enters the official system. The output is organized, confident, and professionally worded. This is precisely where the risk concentrates, because **fluency is not evidence**. A language model can produce a summary that reads as authoritative while containing no traceable basis for its conclusions. The traditional signal that a record was thin, that it read as rushed or vague, is exactly what a language model smooths away. The result is a record that looks ready and is, on inspection, a gap.

The consequence is not that AI must be removed from documentation. It is that the documentation layer now requires a review discipline that standard investigative and administrative guidance does not yet provide.

1.4 What JRS is, and what it is not

JRS is a documentation-review standard. It asks one question of a record: *can a later reviewer reconstruct how a conclusion was reached from the record alone?* It evaluates the record, not the decision, and not the technology that produced the record.

JRS is **not** a legal-compliance framework, a records-retention program, a model-risk-management methodology, an AI-governance framework in the conventional sense, or a substitute for statutory obligations. It does not assess whether an underlying decision was substantively correct. It does not establish legal sufficiency. These boundaries are not incidental; they are load-bearing, because they are what make the standard defensible and honest about its own scope.

---

2. Conceptual Foundations

2.1 Documentation quality versus documentation defensibility

Conventional documentation review asks whether a record is complete, well-written, and policy-compliant. JRS reframes the question around **defensibility**: whether the record contains the identifiable evidence, traceable reasoning, and accountable decision path that an independent reviewer would need to evaluate it without external explanation. A record can be high-quality by conventional measures, fluent, thorough, and neatly formatted, and still be indefensible, because none of those properties guarantees that the basis for its conclusions is present on the page.

2.2 The author-blind principle

A central and deliberately counterintuitive commitment of JRS is that it is **author-blind**. It does not ask, and does not attempt to detect, whether a record was written by a human or by an AI system. Attempting to detect AI authorship is both technically futile and conceptually misdirected. A human-authored record can be conclusory and unsupported; an AI-assisted record can be fully traceable. What matters is not the origin of the language but whether the reasoning survives separation from the person who produced it. For AI-assisted work specifically, this is the correct and durable criterion, because it does not depend on identifying a moving technological target.

2.3 Reconstructability as the core criterion

The unifying criterion beneath all five conditions is **reconstructability**: the property that a neutral reviewer can rebuild the logic of a decision from the record's own contents. Reconstructability is testable, teachable, and independent of both author and technology. It converts an abstract concern about "documentation quality" into a concrete, answerable question that can be applied consistently across domains and record types.

2.4 The proportionality principle

Not every record requires the same depth of anchoring. The documentation defensibility a record must carry is proportional to the stakes of the decision it supports: the tolerable level of Decision Reconstruction Risk falls as the consequence of the decision rises. A transaction that is flagged, reviewed, and closed may be adequately served by a brief action summary. A termination, a compliance red flag, or a clinical determination demands specific, identifiable evidence and, at minimum, references to where the supporting material resides, even when that material is not itself attached. This principle, surfaced by pilot reviewer Saurabh Nanda, General Manager and APAC Business Leader at Align Technology, is what the escalation-routing element of enterprise deployment (Section 8) operationalizes: high-stakes record types route to secondary review precisely because the cost of an unreconstructable record scales with what the decision decides. Proportionality keeps the standard practical. It concentrates review effort where the exposure is greatest rather than imposing uniform documentation weight on records that do not warrant it.

---

3. The Five Conditions

JRS evaluates a record against five conditions. Each is assessed independently. Failure of any single condition identifies a documentation-traceability gap and is a structured trigger for additional review or return to the originating drafter.

Condition I — Reconstructability

A neutral reviewer can identify the basis for the conclusion using only what the record contains, without verbal explanation or outside context. The operational test is whether the record functions with the original author unavailable to explain it. Self-sufficiency is the master condition; the remaining four specify what a self-sufficient record must contain.

Condition II — Basis Identification

Each material conclusion is supported by specific, identifiable evidence: dated records, correspondence, performance or tenancy records, screening or eligibility criteria, documented interactions, or witness accounts. General impressions and unsupported characterizations do not satisfy this condition. An evaluative term such as "unprofessional," "uncooperative," or "unsuitable" requires a documented behavioral anchor regardless of whether a person or a tool introduced the term.

Condition III — Chronology

Relevant events, dates, and their sequence are identifiable from the record. Missing or unidentified dates are, empirically, the most common single deficiency observed in administrative files. A pattern claim (for example, "a pattern of lateness") requires at least two dated, identified instances; a sequence that cannot be reconstructed cannot be evaluated.

Condition IV — Decision-Process Traceability

The path from evidence to conclusion is visible within the record. Conclusions that depend on unstated assumptions or on context absent from the file do not satisfy this condition. This condition distinguishes JRS from documentation-quality checklists that assess only whether a conclusion *appears* reasonable. A coherent narrative is

not equivalent to a traceable record. When Condition IV is absent, a record may contain valid evidence and still fail reconstruction review, because the reasoning that connected the evidence to the conclusion was never committed to the page. The full traceable chain JRS seeks to reconstruct runs:

```
evidence -> AI assistance -> human review -> analysis -> recommendation ->
decision -> authority -> documentation -> independent review
```

Traceability converts *procedural* accountability (a process was followed) into *documentary* accountability (the record itself can show it). Only the second survives a reviewer who was not present.

Condition V — Evidentiary Sufficiency

The evidence in the record is sufficient to support the conclusion drawn. Where automated drafting tools contributed, the source material is identifiable and was reviewed by a human before finalization. AI-assisted wording that introduces characterizations absent from the source notes is a traceability deficiency, not a stylistic improvement.

---

4. The Determination Scale and the Source Credibility Score

4.1 The three-state determination

A reviewer applying JRS renders one of three determinations for a record:

- **Ready** — self-contained; evidence identifiable; reasoning traceable; the basis holds without explanation from the author.
- **Needs work / Review** — the conclusion may be accurate, but the basis is not visible in the record; additional anchoring or secondary review is required.
- **Gap** — the basis for the conclusion is not present in the record; return to the drafter is indicated before the record is finalized.

This three-state scale (rather than a binary pass/fail) preserves the operationally important middle case: records that are probably sound but not yet reconstructable.

4.2 The Source Credibility Score (SCS)

For records making multiple discrete claims, JRS provides a coarse quantitative indicator, the **Source Credibility Score**:

```
SCS = (contemporaneous sources mapped ÷ total claims) × 100
```

The SCS is not a validated psychometric instrument and is not presented as one. It is a triage signal: a low ratio of anchored claims to total claims flags a record for closer review. Its value is directional, not diagnostic.

---

5. AI Documentation Risk

Automated drafting introduces failure modes not present in manually drafted records. JRS catalogues four recurring patterns that reviewers should recognize.

1. **Unsupported narrative generation.** The tool produces a plausible, professionally worded summary not anchored to identifiable source material. The wording appears complete while the evidentiary basis is absent.

2. **Characterization without anchor.** Evaluative language ("resistant to direction," "problem tenant," "unqualified applicant") is introduced by the tool and carries the same anchoring requirement as human-authored language. The

origin of the term does not relax the requirement.

3. **Conflict obscuring.** Automated summarization resolves or omits conflicting accounts, partial recollections, or qualifications present in the source notes. The final record reads as settled while the underlying material is contested or incomplete.

4. **Cross-record wording carry-forward.** Where a tool references prior records during drafting, unanchored characterizations propagate forward. Each repetition appears to corroborate the last, though no individual entry contains independent support.

Each pattern is a specific, checkable manifestation of the general failure JRS is built to catch: text that reads as supported without being reconstructable.

---

6. Measurement and Instrumentation

6.1 The Record Interrogation Framework

The operational core of applied JRS is the Record Interrogation Framework: a structured set of paired reviewer questions and traceability checks that convert each condition into a concrete prompt. A statement such as "the employee demonstrated a pattern of unprofessional conduct" triggers the questions "what specific conduct, on what dates, and is it documented?" and the check "observable anchor required; pattern claims require at least two dated, identified instances." The framework has been instantiated for employment (EEO), fair-housing, and jurisdiction-neutral international investigative contexts.

6.2 Submission readiness triage

At the point of finalization, JRS resolves to a three-state readiness signal, STOP (return to drafter), REVIEW (clarification required), READY (self-contained), that maps directly onto the determination scale and routes elevated-risk records to secondary review before they enter the system of record.

6.3 Operational materials

Enterprise application is supported by a small set of shared instruments: a Pre-Finalization Worksheet, a Rapid Review Card, a Secondary Review Escalation Form, and an AI Verification Checklist. These are designed to insert into existing review steps rather than to constitute a new workflow.

---

7. The Evidence Development Program

JRS is in an operational validation phase. It makes no claim of proven effectiveness. The purpose of the Evidence Development Program is not to assert that JRS works, but to observe, honestly and in public, how reviewers apply the conditions, how consistently they agree, and how evidence accumulates. The program is organized as a three-rung ladder, deliberately distinguishing three properties that are often conflated.

7.1 The three-rung validation ladder

- **Rung 1 — Reproducibility.** Whether independent systems, applying the standard to the same constructed records, agree. The current automated observation is a cross-vendor reproducibility check in which three independent AI models from different vendors each judge the same set of constructed (synthetic) records. Cross-vendor agreement is a stronger signal than a single model repeating itself, but agreement is not accuracy and not validation.
- **Rung 2 — Accuracy.** Whether reviews correspond to an expert benchmark. This rung uses a **reference-panel design**: rather than scoring against a single presumed-correct "gold" answer, review outputs are evaluated against the distribution of judgments from a panel of qualified domain experts labeling a common record set. The

reference-panel framing, contributed by the program's model-validation methodology (see Acknowledgments), treats expert consensus as an empirically derived reference standard and preserves, rather than discards, the informative disagreement among experts.

- **Rung 3 — Criterion validity.** Whether JRS reads correspond to real-world outcomes: de-identified public determinations paired with their documented dispositions (upheld, overturned, challenged), to test whether a JRS read predicts how a record holds up when it is contested.

The three rungs are kept strictly separate. Reproducibility, accuracy, and validity are distinct properties and are not treated as equivalent anywhere in the program.

7.2 Study registry and the AI-Assisted Records Detection pilot

The program maintains an open registry of studies spanning reproducibility, ground-truth benchmarking, per-condition performance, framework ambiguity, reviewer recognition, and criterion validity. A representative study, the AI-Assisted Records Detection pilot, presents reviewers with constructed records, half grounded in identifiable source evidence and half fluent but unsupported, and asks them to judge each blind against a held-out key. The pilot tests whether a trained reviewer, applying JRS, can distinguish a genuinely grounded record from one that only reads as grounded.

7.3 Methodological commitments and honest status

Two methodological commitments govern the program. First, **no fabricated results**: constructed stimuli with known ground truth are used freely and are labeled as constructed, but no scores are invented and no preliminary figure is presented as a validated finding. Second, **appropriate reliability statistics**: because determinations are expected to be skewed toward particular categories, inter-rater agreement is assessed with a framework that reports observed agreement alongside multiple chance-corrected coefficients, Krippendorff's alpha, Fleiss' kappa, and Gwet's AC1, and interprets them together rather than relying on any single number. This framework is deliberately built to withstand the **kappa paradox**, in which high observed agreement coexists with a near-zero kappa under skewed marginals; Gwet's AC1 is reported precisely because it is robust to that condition, where kappa is not. Reliability is evaluated against a **pre-registered analysis plan** specifying, in advance, the agreement thresholds ("floors") below which a reliability claim will not be made, so that the standard for a positive result is fixed before the data are seen rather than chosen afterward. The reference-panel design and this reliability framework are methodological contributions of the program's model-validation adviser (see Acknowledgments). A claim of usefulness would be supported only if independent reviewers, applying the five conditions to records they did not author, identify deficiencies that standard review misses and agree with one another above chance; it would be weakened or falsified if reviewers cannot apply the conditions consistently, if flagged records prove no less reviewable than unflagged ones under expert assessment, or if agreement is no better than chance. As of this writing, the accuracy and reliability rungs are still accruing data, which is why nothing in the program is presented as validated.

7.4 Pilots conducted to date (interim status)

Four lines of work are active. Every figure below is preliminary, drawn from a small sample, and is reported to describe the program's status, not to assert a finding.

Rung 1 — Cross-vendor reproducibility (constructed records). Three large language models from three independent vendors were each applied to a common set of constructed records under the JRS conditions. Across a synthetic set, the models' determinations agreed at approximately 90 percent. This measures only whether independent systems agree on constructed records. It is not accuracy, it involves no human reviewers and no real-world outcomes, and it does not bear on validation. Cross-vendor agreement is nonetheless a stronger signal than a single model repeating itself.

Rung 2 — Expert reference panel (bench review). Constructed records were assembled into batches and reviewed by a panel of domain experts spanning human resources, compliance and data privacy, records and information governance, AI governance and assurance, and general management, together with a larger group of non-expert reviewers. For one fully covered batch, three experts independently reviewed the same five records: observed pairwise agreement was approximately 73 percent, with unanimous agreement on three of the five records. Under the skewed determination marginals of that batch, Fleiss' kappa fell near 0.17 while Gwet's AC1 was approximately 0.68, a direct empirical illustration of the kappa paradox and of why the program reports AC1 alongside kappa. With three experts and five records, these values are interim and indicative only; they are not a reliability result.

AI-Assisted Records Detection pilot. A set of 24 constructed records, half grounded in identifiable source evidence and half fluent but unsupported, was prepared with a held-out key. The first reviewer to complete the full set classified records consistently with the intended design in roughly four of five cases, correctly passing most grounded records and flagging most unsupported ones. A small number of confident but unsupported records were marked ready, the specific failure mode the pilot exists to surface. As a single-reviewer result this is an anecdote, not evidence; additional qualified reviewers are completing the same set so that agreement and above-chance detection can be assessed.

Rung 3 — Criterion validity (public determinations). Collection of de-identified public determinations paired with their documented outcomes is underway across public-records and employment domains, contributed by domain practitioners. The sample is currently too small for analysis, and no outcome correspondence is reported.

Interim caveat. None of the above constitutes a validated finding. Sample sizes are small, several lines rest on constructed rather than real records, and the accuracy and reliability rungs are still accruing data. The pilots demonstrate that the instrument runs and produces clean, structured data; they do not yet demonstrate that it works. That distinction is maintained deliberately.

---

8. The Enterprise Plan

JRS enterprise implementation is a lightweight review layer inserted into existing documentation workflows. It is not a governance-system replacement, a software platform, or an organizational restructuring. Its design principle is that the review is added to the step where review already occurs, before records enter official systems.

8.1 What enterprise application means

Six elements define an enterprise deployment:

1. **Workflow insertion.** Review controls are placed at the existing pre-finalization step. No new workflow stage is created; the question set is added to a step that already exists.

2. **Reviewer onboarding.** Reviewers across departments work from common materials (the Pre-Finalization Worksheet, the Rapid Review Card, the Secondary Review Escalation Form) and can be onboarded using existing training infrastructure.

3. **Phased record-type rollout.** Most organizations begin with one or two high-exposure record types, typically terminations and formal discipline, and expand based on where documentation failures have historically surfaced.

4. **Cross-department standardization.** Review language, escalation routing, and secondary-review protocols become consistent across HR, compliance, and investigation functions, using common standards rather than a dedicated platform.

5. **Escalation routing.** High-risk records (terminations, accommodations, AI-assisted documentation) route to secondary review before system entry, with routing criteria applied consistently across departments.

6. **AI-assisted documentation controls.** Where AI drafting tools are used, the AI Verification Checklist provides a consistent pre-submission control: source verification, human-reviewer confirmation, and no unverified characterizations.

8.2 Where scrutiny is shifting

The enterprise case rests on an observable shift in external review. Reviewers increasingly ask not only what was decided but how, what the AI contributed, what a human verified, and whether the process can be reconstructed from the record. Public-records exposure is widening to reach AI artifacts (prompts, drafts, recommendations, human-review logs). Model governance and records management, historically separate disciplines, are converging because AI-generated content becomes permanent record. The emphasis is moving from documentation *quality* to documentation *defensibility*. Because the same review supports records prepared for litigation, investigations, audits, public-records requests, and regulatory examinations, one control serves many governance functions. And because JRS evaluates the record rather than the technology, the control is durable across changes of model, vendor, or platform.

8.3 Sector applications

The same control applies wherever AI-assisted work becomes a record that must later be reconstructed and defended:

- **Financial services.** Underwriting memoranda, suspicious-activity narratives, alert dispositions, investment research, and audit reports that surface in examinations, enforcement, and lookbacks.
- **Healthcare.** Clinical notes, discharge summaries, prior-authorization and medical-necessity determinations, and peer-review documentation examined in malpractice, payer audits, and accreditation, often years after care.
- **Human resources.** Investigation summaries, disciplinary memoranda, performance reviews, accommodation analyses, and terminations that reappear in agency charges, arbitration, and litigation.
- **Research and knowledge work.** Literature reviews and analytical reports that blend source evidence, AI synthesis, and human interpretation, and that later inform policy, compliance, or investment decisions.
- **Public sector and records.** Determinations and AI artifacts reachable through public-records requests, where the operative question is whether the agency can explain how each record was created.

8.4 Deployment sequence, scenarios, and audit sampling

Enterprise deployments typically progress from selective introduction to broader operation: an initial engagement on a narrow record type, reviewer onboarding via the deployment kit, establishment of secondary-review routing, calibration across the department, and expansion by record type. Common scenarios include a full-department rollout, a simulation-based onboarding pilot in which reviewers calibrate on training exercises before applying controls to live records, and a phased high-risk-first rollout. At scale, the documentation failure-mode catalogue supports periodic **audit sampling**, applying the conditions to a sample of finalized records to identify the record types and workflow stages where documentation quality is consistently insufficient.

8.5 Complementary positioning and enterprise risk reduction

JRS does not replace governance, risk, and compliance platforms, records-management programs, model-risk-management systems, or responsible-AI programs. It evaluates the documentation flowing through them, which makes it relatively low-friction to adopt and additive to existing investment. The enterprise value is risk reduction, consistency, and defensibility: fewer adverse findings where a record's basis is not reconstructable, more consistent determinations across staff, and improved audit and oversight readiness, delivered through one documentation control that spans multiple functions.

8.6 Reviewer calibration through the simulation training platform

The consistency on which JRS depends is not assumed; it is trained. A self-paced simulation training platform prepares reviewers to apply the five conditions before they act on live records, and is the mechanism behind the simulation-based onboarding scenario described above. Its design has five features.

1. **Modular structure.** Six sequential modules carry a reviewer from the conceptual basis of reconstructability, through each of the five conditions, to full record review. Completion of each module is tracked and persisted locally, so training can be paused and resumed and a reviewer's progress is recoverable.

2. **Role-gated paths.** On entry, a reviewer selects a functional role, human resources, compliance, investigations, employee relations, or administration, and the platform renders a path calibrated to the record types that role encounters. The five conditions are taught uniformly throughout; only the illustrations differ by domain, which preserves a single standard while making the training concrete for each reviewer.

3. **Simulation exercises.** Rather than passive reading, reviewers practise on constructed records through targeted exercises: escalation review (deciding when a record should route to secondary review), chronology reconstruction (rebuilding a sequence of events from a record's own dates), and AI-interpretation review (identifying unsupported characterizations introduced by drafting tools). Because the stimuli are constructed, a trainee's judgment can be compared against an intended answer without exposing any real record.

4. **Calibration survey.** A structured survey captures the reviewer's judgments and self-assessment across the exercises, providing calibration data and feedback and contributing to the program's understanding of reviewer behavior.

5. **Completion and recognition.** A reviewer who completes the modules receives a certificate of completion, marking readiness to apply controls to live records.

The training platform is more than onboarding. It is a direct lever on reliability. By calibrating reviewers on a common standard before live use, it reduces avoidable inter-reviewer variance, so that the agreement measured on the accuracy rung reflects the instrument itself rather than uneven training. In this sense the training platform and the Evidence Development Program are complementary: one seeks to remove noise that the other is designed to measure.

---

9. Domain Applications: Administrative Investigations

The investigative setting where JRS originated remains its clearest application. Field guides instantiate the standard for three contexts: employment (EEO/EEOC intake and workplace investigations), fair housing (tenant-screening, availability, and reasonable-accommodation determinations), and a jurisdiction-neutral international edition for equality bodies and national human rights institutions, most of which carry a combined employment-and-housing mandate. In each, the five conditions and the Record Interrogation Framework are applied at intake and before finalization, with elevated-risk records routed to secondary review. The method is jurisdiction-neutral by construction; only the illustrative statutory references change.

---

10. Limitations and Ethical Boundaries

JRS is bounded deliberately, and stating those bounds precisely is itself an application of the standard.

- **It is not legal advice and does not establish legal sufficiency.** Organizations must engage qualified counsel on questions of legal sufficiency, regulatory compliance, and jurisdiction-specific requirements.
- **It is in a validation phase and makes no proven-effectiveness claims.** Its value proposition is risk reduction and consistency, not a guarantee against adverse outcomes.
- **Documentation quality is not decision quality.** A record can satisfy all five conditions and still reflect a decision that is contestable on substantive grounds. JRS evaluates the file, not the outcome.

- **It is administratively neutral.** It assesses whether the record contains the documented basis for its conclusions, not whether those conclusions are correct.

These limitations are the source of the standard's credibility. A record-level control whose entire premise is defensibility cannot afford to overstate its own.

---

11. Strategic Positioning and Intellectual-Property Considerations

Viewed strategically, JRS occupies a governance layer situated between AI-assisted operations and organizational accountability. It evaluates whether AI-assisted documentation remains reconstructable, understandable, evidentially grounded, reviewable, and defensible before it becomes permanent organizational evidence. Three properties support the durability of the intellectual property. It is **technology-agnostic**: it evaluates records, not systems, and therefore does not expire when a model or vendor is replaced. It is **cross-functional**: the same methodology applies across legal, compliance, audit, HR, healthcare, investigations, and records management. And it is **complementary**: it strengthens existing governance investments rather than competing with them. Distribution of the standard as an open reference, with authorship and mark preserved on every instrument, is intended to build recognition and adoption, the network effects that convert a documented method into an institutional standard.

---

12. Conclusion

The significance of the current moment is not that AI has introduced a new kind of record. It is that AI has made an old problem, the gap between when a decision is made and when it is judged, far more common and far harder to see. Fluent, machine-assisted documentation can read as finished while lacking the traceable basis that later scrutiny requires. JRS responds with a single, disciplined question applied before a record becomes permanent: can an independent reviewer reconstruct how the conclusion was reached from the record alone? Its five conditions operationalize that question; its Evidence Development Program tests it honestly and in the open; its enterprise architecture inserts it into existing workflows at low cost. JRS does not govern the model and does not judge the decision. It governs the record, which, when the people and the context are gone, is the only thing that remains.

Acknowledgments

The validation methodology described in Section 7 reflects the methodological contributions of **Ubayet Hossain, FRM, Associate Director (Model Validation), KPMG India**, who developed the reference-panel design for the accuracy rung and the chance-corrected inter-rater reliability framework, including the use of Krippendorff's alpha, Fleiss' kappa, and Gwet's AC1 together, the explicit treatment of the kappa paradox under skewed marginals, and the pre-registered analysis plan with predefined agreement floors. These contributions establish the standard of evidentiary rigor to which the Evidence Development Program holds itself. The proportionality principle (Section 2.4) was surfaced by **Saurabh Nanda**, General Manager and APAC Business Leader (Align Technology), during the validation pilots, and is credited with his permission. Responsibility for the framework as presented, and for any error in its description, rests with the author.
