

EXAMPLE AI TRANSFORMATION DIAGNOSTIC DECISION PACK

Cedar Ridge Benefits Services

A fictional regulated-services case demonstrating how Bravura diagnoses the work first, separates conventional process repair from genuine AI opportunity, and gives leadership a decision it can act on.

OUR UNDERSTANDING OF THE CLIENT PROBLEM

1. REQUESTS ENTER INCONSISTENTLY

Change requests arrive through multiple channels with missing or ambiguous information. Analysts spend time repairing intake before they can do the work the client actually values.

2. EVIDENCE HAS TO BE REBUILT REPEATEDLY

Analysts search for governing material, reconcile versions, and reconstruct context case by case. That creates delay, rework, and avoidable dependence on experienced people who know where the answers usually live.

3. EXCEPTIONS CONCENTRATE WORK IN SCARCE PEOPLE

Ambiguous cases, source conflicts, and policy questions leave the routine queue and land with managers or subject-matter experts. Faster intake can worsen that bottleneck if the operating model does not change with it.

4. THE ORGANIZATION DOES NOT YET KNOW WHAT AI TRULY ADDS

Some burden can likely be removed with better forms, workflow design, and explicit rules. AI should have to prove that it adds useful value beyond those simpler repairs without shifting hidden review work or decision risk downstream.

THE DECISION LEADERSHIP NEEDS TO MAKE

What is the smallest, safest experiment that can show whether AI adds enough incremental value to justify its cost and control burden? The decision is not whether Cedar Ridge should “adopt AI.” It is what to test, which decisions stay with named owners, and what evidence must return before anything expands.

BRAVURA RECOMMENDATION

Proceed with conditions. Run one six-week Workflow A experiment. Improve process and rules first, let AI compete only for the remaining unstructured work, keep consequential decisions with named people, and leave Workflow B closed until source-conflict and decision-ownership rules are explicit and tested.

THE DIAGNOSTIC MUST ANSWER THREE THINGS BEFORE ANYTHING EXPANDS

1. WHAT DOES SIMPLER REPAIR REMOVE?

Measure how much avoidable burden disappears after better intake, required fields, routing rules, and source control. AI should not receive credit for value the simpler repair creates.

2. WHAT DOES AI ADD AFTER THAT?

Test whether AI materially reduces the remaining extraction and retrieval burden while preserving quality, source traceability, and clear decision ownership.

3. CAN THE OPERATING MODEL SUSTAIN IT?

Count the full review and control burden, prove safe fallback and recovery, and require the economics to remain credible before scope increases.

Illustrative-example boundary: every organization detail, count, metric, interview, source item, threshold, cost assumption, and scenario in this example is synthetic. The specificity demonstrates consulting reasoning and delivery quality; it does not represent an actual client result.

Authorize one six-week experiment before making a broader AI commitment.

The immediate decision is what to test, which decisions stay with named owners, and what evidence must return before anything expands.

Our recommendation: Authorize a six-week Workflow A experiment. Clean up intake and routing first, then compare that rules-only design with AI assistance on the same cases. Keep final classification, source conflicts, exceptions, policy interpretation, risk acceptance, and external commitments with named owners. The experiment should tell leadership which work is best handled by process design, conventional automation, AI assistance, or expert judgment before any broader commitment is made.

Cedar Ridge's problem is not "how to deploy AI." It is how to reduce reconstruction work without letting a machine make decisions whose ownership and rules are still unresolved. Requests arrive in inconsistent forms. Analysts repair them, search across sources whose applicability can be conditional, classify the work, and escalate exceptions. The same operating defect that creates delay also makes broad generative assistance unsafe: the organization has not yet made enough of its decision-ownership and source-control rules explicit.

The recommendation therefore separates three interventions that are easy to blur together. Stable completeness requirements belong in structured intake and explicit rules. Unstructured field extraction and retrieval of potentially relevant sources can be tested with AI because they can remain evidence-producing rather than decision-making. Final classification, source applicability, exception judgment, risk acceptance, policy interpretation, and external commitment stay owned by named people.

The pilot has to answer whether AI adds incremental value beyond the non-AI repair. If rules and better intake remove most of the burden, Bravura should say so. If AI saves analyst time but creates manager review, verification, or correction work, that burden counts against the result. A successful Diagnostic can therefore return "use less AI," "redesign first," or "stop."

AUTHORIZE NOW

Six-week limited Workflow A experiment with a rules-only comparison, mandatory review by a named decision owner, source traceability, predeclared stop conditions, and tested manual fallback.

DO NOT AUTHORIZE

Generative Workflow B interpretation, autonomous response, exception approval, risk acceptance, AI-created precedent, or scale based on usage/cycle time alone.

RETURN DECISION

Scale within the proven limits, redesign and retest, hold for evidence, or stop. Workflow A success does not automatically reopen Workflow B.

CONFIDENCE

Moderate. The synthetic case supports a learning experiment, not autonomy or scale. AI error classes and net verification burden remain intentionally unknown.

STATUS QUO CONSEQUENCE

Current work remains viable but keeps reconstruction cost and unwritten decision rules hidden; it produces no evidence for safe delegation.

WHAT CHANGES THE RECOMMENDATION

Rules-only captures most value; manager burden rises; source conflicts dominate; quality or recovery fails; or AI-assisted design adds no material incremental benefit.

REVERSIBILITY

High by design. Assisted steps can be disabled by case type and work returns to a tested manual path without changing external decision ownership.

Executive decision chain: evidence sufficiency -> causal alternatives -> solution selection -> limited, reversible experiment -> critical-failure return gate.

What this Diagnostic will answer, and where another owner must take over.

The work is deep enough to support the next decision without pretending to be implementation, procurement, legal advice, or a full compliance review.

What this engagement covers: This Diagnostic goes deep on the two selected workflows and the decisions around them. It does not substitute for implementation, procurement, legal advice, privacy or security assessment, or enterprise-wide AI strategy. That boundary keeps the work accountable: Bravura owns the diagnostic conclusions inside scope and identifies where another owner or specialist must take over.

In scope: Workflow A from receipt through analyst triage and routing; task-level solution selection among process, explicit rules, conventional automation, AI assistance, and role-owned work; governed source retrieval needed by that workflow; verification burden; decision ownership; manager/workforce consequences; pilot measurement; failure recovery; transfer.

Adjacent, but intentionally limited: Workflow B is analyzed only far enough to establish why generative assistance is deferred and what prerequisites would justify reopening that decision.

Out of scope: enterprise AI strategy, production implementation, vendor procurement, legal or regulatory opinion, full cybersecurity/privacy audit, headcount action, and autonomous external communication.

This boundary matters because "AI transformation" can become an excuse to claim expertise over every adjacent domain. Here, any legal or regulatory applicability question that would control a case is a stop-and-escalate condition, not an inference opportunity.

Decision discipline: absent evidence is not neutral space. If a missing quantity could change the recommendation, it becomes part of the pilot or a reason to hold the decision.

CAN DECIDE NOW

Do not authorize broad autonomy. Standardize structured intake first. Keep consequential source conflict, exception approval, risk acceptance, and external commitments owned by named people.

MUST LEARN IN PILOT

Whether AI-assisted design adds value beyond rules-only; how burden moves by role; error distribution by case type; reviewer challenge quality; recovery performance; conservative economics.

REQUIRES SEPARATE OWNER OR APPROVAL

Workflow B interpretation, legal/regulatory applicability, enterprise scale, vendor procurement, headcount action, and any expansion of external decision ownership.

EXECUTIVE DECISION QUESTION

For client change-request work, what should change with process/rules, what can safely receive AI assistance, what remains owned by named people, and what limited, reversible experiment would create enough evidence for the next delegation decision?

EVIDENCE BURDEN

Consequential claims must show support, what the support permits, alternatives, falsifiers, uncertainty, owner, measurement, recovery, and transfer. The synthetic case can demonstrate that reasoning shape; it cannot turn itself into empirical proof.

KEY UNRESOLVED QUANTITIES

AI error distribution by case type; net verification burden; manager queue response; incremental value over rules-only; source metadata quality; automation-complacency behavior; real annual volume and economics.

The evidence is strong enough for an experiment and too early for transformation claims.

Ten synthetic evidence objects support a limited, reversible experiment while leaving the most consequential unknowns visible.

What the evidence supports: The synthetic evidence supports a controlled experiment. It does not support claims of realized savings, enterprise readiness, or safe autonomy. We know enough to compare rules-only and AI-assisted designs, measure burden by case type, and test failure and recovery. The live experiment must still establish the real error mix, review burden, and economics.

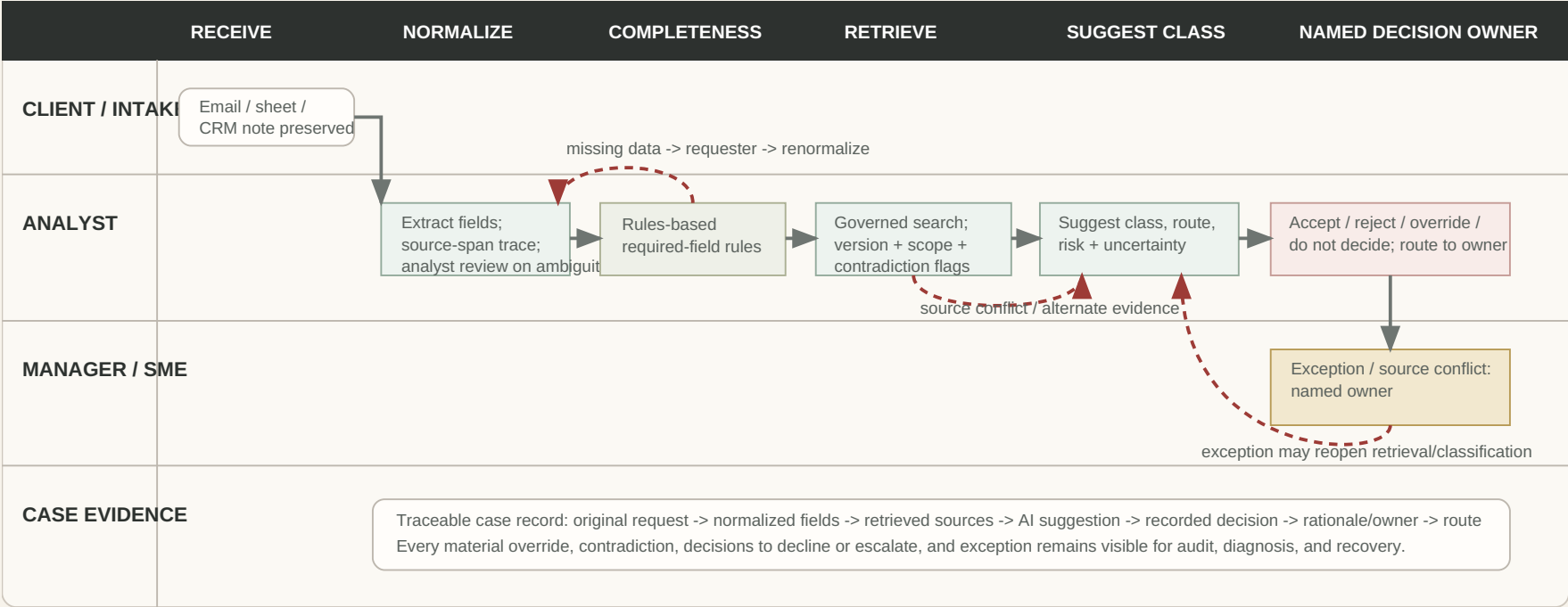
ID / CLASS	SYNTHETIC BASIS	WHAT IT SUPPORTS	WHAT IT DOES NOT PROVE	CONFIDENCE USE
E01 Synthetic Case Trace	24 modeled request traces across routine, complex, ambiguous, and exception-bearing cases	rework loops; handoffs; evidence reconstruction; exception paths	AI effectiveness; causal share of each burden; future ROI	MODERATE FOR STRUCTURE
E02 Synthetic Source Review	11 modeled source items spanning client-specific commitments, policy, approved procedure, knowledge guidance, and prior-case material	source-class fragmentation; effective-date/version need; precedent risk	legal hierarchy; regulatory compliance	MODERATE HIGH WITHIN SYNTHETIC CASE
E03 Synthetic Workflow Walkthrough	Two end-to-end modeled walkthroughs, one routine and one exception-bearing	step sequence; queue and escalation points; hidden coordination work	frequency of each failure; population-wide burden	MODERATE
E04 Synthetic Time Allocation	Illustrative analyst effort decomposition across modeled work	21% source-locating/reconciliation burden; candidate burden categories	actual labor savings; causal impact of AI	MODERATE LOW FOR MAGNITUDE
E05 Synthetic Role Interview	Six fictional role interviews representing analyst, manager/ SME, operations, risk/compliance, client-facing, and tool/data perspectives	decision-ownership inconsistency; manager rescue burden; challenge/escalation needs	workforce sentiment distribution; adoption level	MODERATE FOR HYPOTHESIS
E06 Synthetic Control Review	Modeled review of approval, exception, evidence, change, and external-commitment controls	missing traceable record; need for explicit stop/resume; source-control dependency	compliance deficiency; audit failure	MODERATE HIGH WITHIN SYNTHETIC CASE
E07 Synthetic Metric Sample	50 modeled Workflow A requests used only to illustrate diagnostic measurement logic	follow-up burden; case-mix segmentation requirement; exception cycle spread	root cause; AI opportunity magnitude	MODERATE FOR ILLUSTRATION
E08 Synthetic Case Record Traceability Review	24 modeled case records checked for a complete chain from original request through sources, rationale, decision, approval, and final action	case-record traceability gap	regulatory noncompliance	HIGH WITHIN SYNTHETIC CASE
E09 Synthetic Queue Observation	Modeled queue decomposition separating analyst touch time from manager/SME waiting time	need to separate processing from waiting; manager bottleneck alternative	exact waiting-time share	LOW MODERATE
E10 Synthetic Solution Retest Design	Same-case comparison design comparing current-state, rules/ forms-only, and AI-assisted rules+AI on the same fixed case set	ability to distinguish AI value from process standardization	future live behavior	DESIGN EVIDENCE

Sufficiency conclusion: enough to select Workflow A as a learning surface, require a rules-only same-case comparison, keep consequential decisions with people, and define a pilot. Not enough for autonomy, enterprise scale, headcount action, realized ROI, legal conclusion, or a claim that AI outperforms conventional redesign.

The delay comes from repeated context rebuilding, not one slow step.

Workflow A shows where requests are repaired, sources are reconstructed, decisions are reworked, and exceptions loop back through the system.

Our diagnosis: Most delay comes from rebuilding context: repairing incomplete requests, searching and reconciling sources, then handing exceptions to scarce managers or specialists. Speeding up one step can simply move the queue. Bravura would instrument those loops first, remove preventable rework with better intake and rules, and then test whether AI reduces the remaining search and extraction burden without creating a new review bottleneck.



LOOP 1 - MISSING DATA

Incomplete or ambiguous requests return to the requester. The illustrative 34% follow-up signal makes completeness a measurable operating burden, not a vague frustration.

LOOP 2 - SOURCE RECONSTRUCTION

Analysts search and reconcile possible sources. Retrieval can help only if source identity, version, scope, and conflict travel with the result.

LOOP 3 - EXCEPTION OWNERSHIP

Ambiguous cases leave the analyst queue for a named decision owner. Upstream speed can make this queue worse, so manager wait and touch time must be measured separately.

WHAT WE SEE

The rework loops are the operating problem. The process repeatedly recreates context before judgment can happen.

WHAT WE WOULD CHANGE FIRST

Standardize intake, required fields, source identity, and obvious routing rules before asking AI to absorb variability.

WHAT AI MUST EARN

A measurable reduction in remaining search/extraction burden without increasing manager review, corrections, or exception queues.

Time saved for analysts can become hidden work for managers.

Touch time, queue time, verification, and correction must be separated so apparent productivity does not conceal a new review burden.

What leadership should watch: Analyst minutes are only one part of the cost. If managers or subject-matter experts inherit more checking, exception work, or correction, the system has not improved. The pilot should track analyst, manager, specialist, correction, queue, and escalation effort together so a faster front end cannot hide a slower operating model.

Touch time is active analyst or manager work. **Queue time** is waiting for a requester, source owner, manager, or SME. **Correction time** is work created because a prior step was wrong or incomplete. Those three mechanisms respond differently to intervention.

A structured intake rule can reduce avoidable follow-up. Controlled source retrieval can reduce search touch time. Neither necessarily reduces manager queue time for genuinely complex exceptions. An AI suggestion can even increase correction or verification time while making the analyst step appear faster.

That is why the pilot's economic denominator is not "AI minutes saved." It is **net work minutes per correctly resolved case**, segmented by case type and role. End-to-end cycle time remains useful, but it cannot tell us where the burden went.

Measurement implication: timestamp step entry/exit, record active work minutes by role, record follow-up loops and source conflicts, and preserve downstream correction. Without that measurement setup, the pilot cannot distinguish acceleration from burden transfer.

PATTERN A · REAL PROCESS GAIN

Follow-up loops and analyst touch fall in both rules-only and AI-assisted conditions, while manager burden stays flat. Conclusion: standardization created value; AI receives credit only for any additional lift over rules-only.

PATTERN B · BURDEN MIGRATION

Analyst touch falls but manager/SME minutes, corrections, or exception waits rise. Conclusion: front-end speed is not net productivity. Narrow, redesign, or stop until the displaced burden is addressed.

PATTERN C · HIDDEN COMPLEXITY EXPOSED

Recorded exceptions rise while wrong decisions fall. Conclusion: the old process was suppressing ambiguity. Higher visible exception volume can be healthy if ownership and capacity are sufficient.

TIMESTAMP THE WORK

Record step entry and exit so queue delay is not confused with active work.

COUNT WORK MINUTES BY ROLE

Separate analyst, manager, SME, correction, and escalation time rather than reporting one blended saving.

RECORD REWORK AND OVERRIDE

Count follow-up loops, corrections, rejected suggestions, overrides, and reopened cases.

TRACK QUEUES AND SOURCE CONFLICT

Measure manager wait, exception backlog, and source-conflict cases so acceleration does not hide a new bottleneck.

34%

illustrative follow-up signal. It suggests an intake problem, but does not prove that AI is the solution.

21%

illustrative analyst effort in source location/reconciliation. It motivates controlled source retrieval, not autonomous interpretation.

2.39x

illustrative median cycle spread between ordinary and exception-heavy cases. Complexity and manager queues remain competing explanations.

0/24

modeled case records with a complete request-to-source-to-rationale-to-approval chain. A missing decision trail is a control defect even before AI.

WHAT WOULD INVALIDATE THE RECONSTRUCTION THESIS?

If case-mix-adjusted data shows that manager waiting or inherently complex judgment dominates total delay, while intake/source reconstruction is small, then the intervention should move downstream or become process-capacity work rather than an AI assistance program.

We kept competing explanations alive before choosing a pilot.

Seven hypotheses make the recommendation challengeable and specify what evidence could weaken it.

Why we kept competing explanations open: Several explanations still fit the evidence: process defects, fragmented sources, unclear decision ownership, a real opportunity for AI assistance, and hidden verification work. The experiment is valuable because it can separate them. A rules-only win, a redesign recommendation, or a decision to stop are all legitimate outcomes.

HYPOTHESIS UNDER TEST	CREDIBLE ALTERNATIVE	WHAT WOULD WEAKEN / DISPROVE IT	CURRENT CONFIDENCE
H1 Incomplete and heterogeneous intake materially drives avoidable follow-up/rekey work.	clients legitimately lack required facts; requirements themselves are unclear	After standard fields/rules are introduced, follow-up rate and minutes do not materially decline after case-mix adjustment.	MODERATE HIGH
H2 Source reconstruction and unclear decision ownership materially drives analyst and exception burden.	search tooling alone is poor; sources are simple but analysts lack training	Controlled source retrieval materially improves search without affecting rework/manager burden, and source conflicts are rare after segmentation.	MODERATE HIGH
H3 Exception complexity, not reconstruction burden, may explain most cycle-time spread.	manager queue; client wait	Within matched case types, completeness/source-reconstruction measures remain strong predictors of cycle time after exception status is controlled.	UNRESOLVED
H4 Analyst skill/tenure variation may drive routing and search differences.	case mix differs by analyst	After stratifying by case type, analyst-level variation is small relative to intake/source effects.	LOW MODERATE
H5 Manager/SME queueing may be the dominant bottleneck in exception cases.	source search occurs before escalation	Manager wait/touch share remains small and does not rise when upstream processing accelerates.	UNRESOLVED
H6 Conventional structured intake and rules may capture most of the attainable value without AI.	process-only redesign; rules/forms-only	On same-case retest, AI-assisted design materially reduces net burden and/or improves evidence retrieval beyond rules/forms while meeting quality and decision-control thresholds.	MODERATE
H7 AI assistance may create hidden downstream verification burden that erases gross time savings.	verification may decline as source trace improves	Net work minutes per correctly resolved case fall with no material manager/SME burden shift and stable quality across case types.	MODERATE

Most decision-relevant unresolved hypothesis: H6. If structured intake and explicit rules capture most of the benefit, AI should not receive credit for process repair. Week 2 therefore uses same-case retest to compare rules-only with the AI-assisted design before live activation.

IF PROCESS REPAIR WINS

Standardize intake and rules, reduce the AI scope, and keep the simpler operating model.

IF SOURCE CONTROL WINS

Fix the source registry, applicability rules, and decision ownership before expanding AI assistance.

IF AI ADDS INCREMENTAL VALUE

Scale only the specific assisted jobs that clear quality, burden, economics, and recovery conditions.

Much of the opportunity is simpler than AI, and that is good news.

Task-level solution selection shows which problems belong to process design, explicit rules, AI assistance, or expert judgment.

Our recommendation: Match each task to the simplest mechanism that can do it reliably. Forms and explicit rules should handle stable requirements; conventional automation should handle repeatable routing; AI should be tested only where language or retrieval variability defeats simpler logic. Analysts, managers, or specialists remain accountable for consequential judgments. This keeps the operating model, rather than the technology, in charge of the design.

WORK OBJECT	BEST FIRST MECHANISM	WHY	DECISION OWNER
Preserve original request / case ID	Conventional system	Record-integrity rules. AI adds no useful ambiguity handling.	SYSTEM RECORD
Required-field completeness	Structured intake + explicit rules	Known requirements should be explicit and testable, not probabilistic.	LOW
Unstructured field extraction	Rules + AI assist	AI earns a role only where incoming prose resists stable mapping; original span remains visible.	ASSIST ONLY
Candidate-source retrieval	Governed AI retrieval	Semantic search can reduce reconstruction burden if source metadata and contradictions travel with results.	REVIEWER VALIDATES
Class / route / risk flag	AI suggestion + named reviewer final	Useful learning surface, but error classes and verification burden are unknown.	NAMED REVIEWER FINAL
Source conflict / exception	Named owner	Ambiguity is the reason to escalate, not an invitation for the model to invent the decision rule.	RESERVED
Policy interpretation / client commitment	Manager / SME only in this pilot	Workflow B lacks a mature source-control model and has external consequence.	PROHIBITED

What this means for the pilot: the rules-only design should carry all stable intake, completeness, and routing logic. The AI-assisted condition should differ only where AI is actually being tested for variable extraction, source retrieval, or classification support. That keeps the comparison interpretable and prevents AI from receiving credit for ordinary workflow redesign.

THE DESIGN TEST

AI must beat the non-AI repair on a decision-relevant dimension. If process standardization and rules deliver equivalent quality and burden reduction at lower verification/governance cost, the correct recommendation is rules-first.

If AI-assisted design wins only on gross analyst speed while manager review rises, it has not won.

WHAT AI IS ALLOWED TO PRODUCE

Candidate values, candidate sources, suggestions, confidence/uncertainty signals, contradiction flags, and evidence that a reviewer can inspect. Fluency does not give the system permission to decide.

Tradeoff: this architecture deliberately keeps more review work than a maximum-automation design. The cost is lower apparent automation; the benefit is that the pilot can actually tell whether delegation is safe and economically useful.

AI cannot safely interpret policy until leadership decides which source wins when sources disagree.

Workflow B stays closed because which source controls depends on applicability, time, scope, conflicts, and rightful expert judgment, not just document relevance.

Leadership decision: Keep Workflow B closed for now. Cedar Ridge first needs an explicit rule for what controls when sources disagree and a named owner for resolving exceptions. Better retrieval can surface relevant material, but it cannot settle effective dates, client-specific commitments, superseded guidance, or novel interpretation on its own. Once those rules are explicit and tested, AI assistance can be reconsidered.

CLIENT-SPECIFIC COMMITMENT

May govern client-specific obligation within authorized scope. Requires client, version, effective date, owner, and scope. Never assumed to override mandatory constraints.

CONTROLLING INTERNAL POLICY

Governs internal rule within stated applicability. Semantic similarity does not establish applicability.

APPROVED PROCEDURE

Defines an execution method and should link to the controlling source. Procedure is not automatically policy.

APPROVED GUIDANCE

Supports interpretation when current and approved. Guidance remains distinguishable from the controlling source.

PRIOR CASE / PRECEDENT

Context for a prior decision. It cannot self-promote into policy and must retain its source/rationale lineage.

INFORMAL LOCAL NOTE

Context only. It may help explain practice but is never the controlling source.

1. Retrieve candidates

Return source class, owner, version, effective date, scope and exact passage.



2. Check applicability

Does this source apply to this client, request class, time, and decision?



3. Conflict?

If plausible sources disagree or applicability is unclear, decline to decide and escalate.



4. Named owner decides

Named owner resolves with rationale; legal/regulatory applicability exits this Diagnostic when required.

Workflow B reopen gate: typed source registry, conditional applicability/precedence rules by request class, effective-date/supersession behavior, conflict ownership, criteria for when to decline and escalate, approved interpretation lineage, and change-triggered recheck must all be operational - not merely documented.

REJECT CONTROLLING USE

Missing owner, version, effective date, scope, or applicability is not a retrieval inconvenience. It makes the candidate ineligible to control the case until the source owner repairs or classifies it. Expired or superseded material remains history, not the default controlling source.

DO NOT DECIDE WHEN SOURCES CONFLICT

When two plausible sources disagree, or applicability depends on unresolved client/legal/regulatory context, the system preserves both candidates and metadata, marks the conflict, and routes to the named decision owner. It does not rank by fluency, similarity, or model confidence.

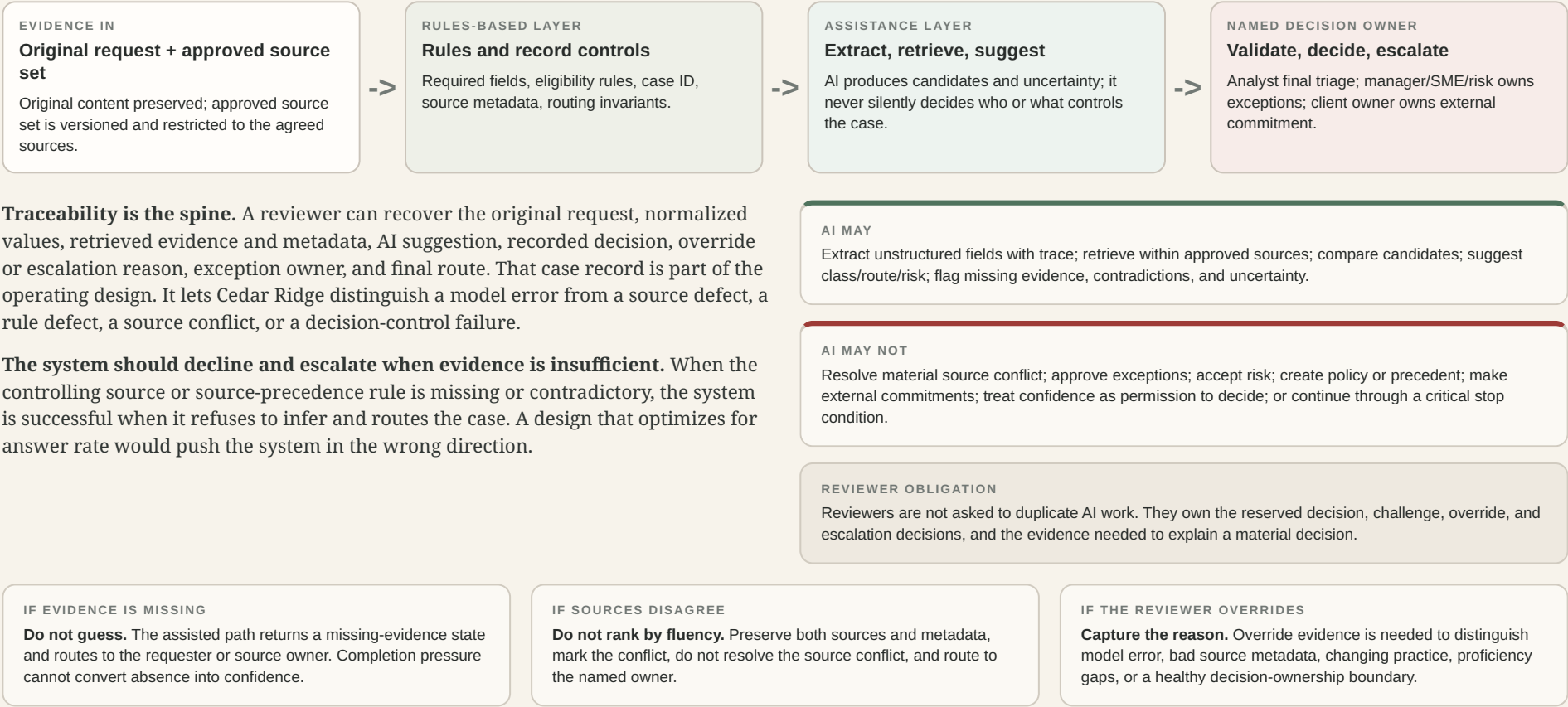
DECISION BECOMES EVIDENCE

The decision owner records the controlling source, rationale, scope, and any exception. That trace is reusable for audit and learning but does not silently manufacture new policy or precedent. Material source changes trigger recheck of affected rules, retrieval, and guidance.

The safest design gives AI specific jobs and keeps consequential decisions with people.

Rules create stable structure, AI produces inspectable candidates, and named owners retain exceptions, risk acceptance, policy interpretation, and external commitments.

Operating principle: Give AI narrow, inspectable jobs. Let it extract, retrieve, and suggest where those tasks add value; let the decision owner accept, reject, override, or escalate. If the system cannot recover the supporting source, express uncertainty, or stay inside its assigned job, it should route the case instead of guessing.



Governance must tell people exactly what happens when a control fails.

Eight controls connect ownership, triggers, evidence, escalation, data/tool boundaries, and stop/recovery behavior to observable operating events.

Governance requirement: Each control needs an owner, a trigger, evidence that it operated, and a defined response when it fails. For this experiment, the priorities are source control, a named final decision owner, approved data and tool paths, exception ownership, and a tested stop-and-recovery process. Abstract principles are not enough if the team cannot execute them under pressure.

CONTROL OBLIGATION	OWNER	TRIGGER	EVIDENCE THAT IT OPERATED	FAILURE / CONTAINMENT BEHAVIOR
G01 Approved source set and registry	Risk/Policy Ops	source add/change/supersession	registry entry + version/effective date + approval state	remove source from retrieval; revalidate affected cases/retest set
G02 Source trace on every assisted retrieval	Operations	every assisted case	source IDs/classes/versions/spans in case record	affected case reverts to manual; repeated failure stops retrieval assist
G03 Named reviewer final decision	Operations manager	every assisted classification	accept/reject/override/escalate decision	contain assisted classification step
G04 Exception ownership matrix	Operations + Risk	exception/contradiction/material uncertainty	owner + response + rationale	pause affected class until ownership resolved
G05 Tool/data-path approval	Tool/Data owner + Risk/Security as required	tool/config/data change	approved configuration and data-boundary record	immediate stop and containment
G06 Versioned prompt/model/retrieval configuration	Tool/Data owner	any config/model/prompt/retrieval change	version + change reason + retest results	no live change until recheck gate
G07 Weekly critical-failure pilot review	Ops sponsor	weekly	scorecard segmented by case type and role	review/contain/stop according to thresholds
G08 Incident stop and recovery decision owner	Risk/Compliance reviewer	critical stop condition	incident record + containment + retest validation	tested manual fallback until restart approval

Non-negotiable decision rule: final Workflow A classification is owned by named people; source conflicts and exceptions route to named owners; external commitments remain with the authorized client-facing owner; the executive sponsor owns scale/redesign/stop.

Change-control invariant: source, prompt, model, retrieval, or tool changes during the pilot are versioned and retested. A silent configuration change destroys interpretability even if output quality appears better.

PRIORITY 1 - SOURCE CONTROL

Know which sources are approved, current, applicable, and superseded. Retrieval without source control creates confident ambiguity.

PRIORITY 2 - DECISION OWNERSHIP

Every assisted classification ends with a named person who can accept, reject, override, or escalate and leave a traceable record.

PRIORITY 3 - STOP AND RECOVER

A control failure must route affected work back to the tested manual process and define exactly what evidence is required before assistance resumes.

The pilot only works if decision-making gets clearer without becoming extra hidden work.

Role changes, manager capacity, proficiency, incentives, challenge behavior, and workload are part of the operating design, not a separate change-management appendix.

Workforce implication: The target work shift is clear: less rekeying and searching, more validation, challenge, and exception resolution. Managers should not become an invisible review queue. Proficiency should be demonstrated on realistic decision patterns, especially source conflicts and escalation, and manager capacity should be measured as part of both the operating and economic case.

Analysts: less rekeying, chasing, and broad search; more evidence validation, exception recognition, challenge/override, and documented reasoning. The failure risk is deskilling if suggestion acceptance replaces source and exception judgment.

Managers and SMEs: less ad hoc rescue is possible only after exception classes and ownership become explicit. Early in the pilot, burden can increase because ambiguity that was previously hidden becomes visible. That is not automatically a defect; it becomes a defect if the burden persists or swamps analyst savings.

Operations leadership: usage and throughput lose their privilege as primary success metrics. Leaders have to govern quality, net operating burden, decision-control behavior, case-mix effects, and recovery evidence together.

Manager capacity is a design constraint. The pilot must budget review/exception capacity. Sending more cases to an already saturated SME queue is not transformation; it is queue relocation.

ENTRY PROFICIENCY GATE

10-case synthetic challenge set before live assisted handling. At least 90% correct decision overall, 100% correct escalation on critical source-conflict cases, and demonstrated ability to override/escalate with rationale. Attendance alone earns nothing.

BEHAVIORAL EVIDENCE

Override with reason; decline or escalation decision; bypass reason; unexplained acceptance; challenge-set performance over time; manager review minutes. These signals distinguish healthy skepticism from model rejection or complacency.

ADOPTION TRAP

High usage can be bad. Appropriate refusal, escalation, and non-use are successful behaviors when the task is outside the allowed scope. Low usage may be resistance or correct judgment; interpretation requires case evidence.

SUSTAINMENT RISK

As assistance improves, independent source and exception judgment can decay. Proficiency refresh and selected manual-review cases preserve the ability to recover when the assisted path is unavailable or wrong.

INCENTIVE DESIGN

Do not reward raw acceptance or usage. Reward correct handling, evidence traceability, appropriate challenge, safe escalation when evidence is insufficient, and timely escalation. Otherwise productivity pressure can train reviewers to rubber-stamp.

CAPACITY TRIGGER

If manager/SME minutes per eligible case enter the review zone or queue age rises materially, narrow the eligible case mix before expanding analyst throughput. Upstream acceleration cannot outrun exception capacity.

DESKILLING DETECTOR

Track challenge-set performance and selected manual-review cases over time. If independent source/exception judgment declines, increase refresh practice and reduce reliance until proficiency is restored.

MANAGER CAPACITY BEFORE LIVE USE

Estimate expected review and exception volume by case type. If the queue cannot absorb it, narrow the pilot before activation.

PROFICIENCY BEFORE PERMISSION

Use realistic challenge cases to verify that people can reject, override, escalate, and recover, not merely complete training.

ESCALATION OWNERSHIP BEFORE SPEED

Name who owns source conflict, exception judgment, risk acceptance, and fallback so faster intake does not create an ownerless queue.

Leadership has five credible operating options.

Status quo, process-and-rules repair, a limited AI-assisted design, broad copilots, and autonomy are all treated as credible options with different learning value and decision-ownership risk.

Our recommendation: Choose the limited AI-assisted design because it creates the most useful learning while preserving a credible rules-only alternative. Broad copilots and autonomous decision paths are premature. If the rules-only design captures most of the value, use the simpler design. That is a successful diagnostic result, not a failure of the engagement.

OPTION	POTENTIAL VALUE	DECISION-OWNERSHIP RISK	LEARNING VALUE	PRINCIPAL TRADEOFF	DECISION
O0 Status quo / no pilot Keep current process and collect better baseline evidence only.	LOW IMMEDIATE	LOW	MODERATE	preserves current burden and delays solution evidence	CREDIBLE ALTERNATIVE
O1 Process + structured intake/rules only Standard fields, explicit completeness rules, route rules, source registry; no AI.	MODERATE	LOW	HIGH AS COMPARISON	may leave unstructured extraction/search burden unresolved	REQUIRED COMPARISON
O2 AI-assisted design with defined limits O1 plus AI extraction of unstructured fields, controlled source retrieval, and classification suggestions under named reviewer final decision.	POTENTIALLY MODERATE HIGH	LIMITED MODERATE	HIGHEST	adds verification/governance burden and new failure classes	RECOMMENDED PILOT
O3 Broad copilot across Workflow A and B Generative assistance extends to policy interpretation and response drafting.	POTENTIALLY HIGH	HIGH UNRESOLVED	LOWER THAN APPEARS BECAUSE HARD TO INTERPRET WHILE CONTROLS ARE UNRESOLVED	speed can amplify ambiguity	DEFER
O4 Autonomous response / exception handling AI interprets and makes external commitments without mandatory named approver.	UNKNOWN	UNACCEPTABLY HIGH FOR CURRENT EVIDENCE	NOT JUSTIFIED	client harm / unapproved decision	DO NOT AUTHORIZE

Why O2: it can tell Cedar Ridge whether semantic assistance adds anything after the rules-based repair. It is reversible, preserves a rules-only comparison, and makes verification and decision-control cost visible. O0 and O1 remain credible if evidence weakens the AI mechanism.

Why not O3/O4: expanding into Workflow B now would mix technical capability with unresolved source-control rules. A poor result would be hard to interpret; a superficially good result could create false confidence. That is low-quality learning at high consequence.

WHY NOT RULES-ONLY NOW?

It may be the right answer. The recommended AI-assisted pilot keeps rules-only as a serious comparison rather than assuming AI must win.

WHY NOT BROAD COPILOTS?

They expand the surface area of review and decision risk before the operating prerequisites are proven.

WHY NOT AUTONOMY?

The current evidence does not justify delegating consequential decisions, source conflicts, exceptions, or external commitments.

CHOOSE RULES-ONLY WHEN

The AI-assisted advantage is small, inconsistent, expensive to verify, or mostly explained by the redesigned intake and rules.

CHOOSE THE LIMITED AI-ASSISTED DESIGN WHEN

AI adds material value after the rules repair and quality, full operating burden, economics, and recovery remain acceptable.

DO NOT ADVANCE WHEN

A broader copilot or autonomous path still depends on unresolved source conflicts, unclear decision ownership, or unsupported external commitments.

Finance should be able to reject the AI path if the economics do not hold.

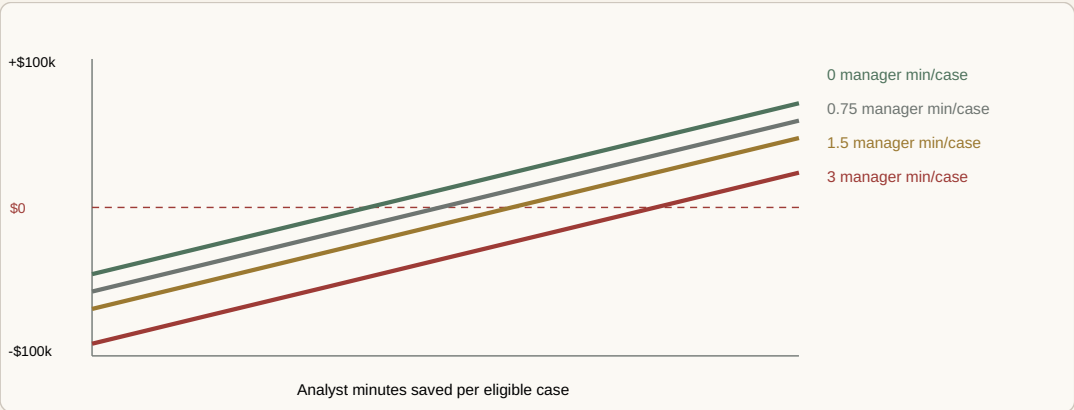
The break-even model makes hidden review burden visible and shows how much net total operating effort must actually disappear before the new operating cost is justified.

What Finance should test: Treat the economics as a rejection test. Under the stated assumptions, the assisted design must remove enough net analyst, manager, specialist, correction, and incident effort to cover the added annual tool and control cost. Manager review moves the break-even line materially. This is a decision threshold, not a savings forecast or a headcount case.

Unit formula: net work minutes per correctly resolved case = analyst touch + manager/SME touch + correction/rework + incident/recovery time.

Annual value formula: eligible cases x net analyst minutes removed x analyst loaded cost - added manager/SME time - incremental tool/control cost - attributable correction/incident cost.

For this public example only, the decision surface assumes 12,000 eligible cases/year, \$48/hour loaded analyst cost, \$78/hour manager/SME cost, and \$45,000 incremental annual tool/control cost. These are synthetic assumptions, not market benchmarks or Cedar Ridge facts.



Lines show synthetic annual net capacity value under different manager-burden assumptions. The zero line is the operating-cost break-even. Actual client economics require actual volume, labor, tool/control cost, correction cost, and a decision about whether capacity can be redeployed.

BREAK-EVEN, FLAT MANAGER BURDEN

4.69 min

analyst minutes saved per eligible case.

BREAK-EVEN, +1.5 MANAGER MIN/ CASE

7.12 min

analyst minutes now required. Hidden review burden materially moves the economics.

ECONOMIC FALSIFIER 1 · RULES-ONLY WINS

If structured intake and rules-based routing remove nearly the same net work minutes as AI-assisted design, the incremental AI mechanism is not economically earned even if the AI-assisted result is positive.

ECONOMIC FALSIFIER 2 · VERIFICATION TAX

If manager/SME or correction minutes rise enough to cross the break-even surface, apparent analyst productivity is consumed by hidden review work. Gross speed cannot rescue the case.

ECONOMIC FALSIFIER 3 · CAPACITY IS NOT CASH

Even a positive capacity estimate is not realized savings unless Cedar Ridge actually redeploys capacity, avoids cost, improves service, or creates another measurable benefit. No headcount action follows from pilot arithmetic.

MEASURE THE FULL CASE

Count analyst, manager, SME, correction, escalation, and ongoing control effort. Partial labor accounting creates false savings.

TREAT BREAK-EVEN AS A TEST

The model tells us what the AI-assisted design must beat. It is designed to make rejection easy if AI cannot clear the burden threshold.

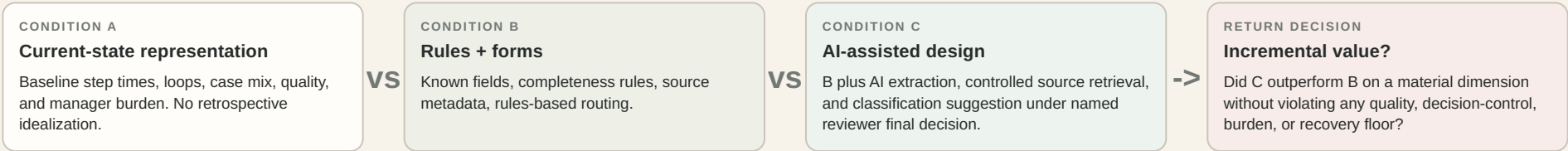
KEEP HEADCOUNT OUT OF THE PILOT CASE

No staffing action is justified until net burden reduction is real, sustained, and separable from temporary learning effects.

The pilot must tell us whether AI adds value beyond a simpler fix.

Using the same cases for the rules-only comparison isolates what AI adds instead of giving AI credit for intake and process redesign.

What the pilot must prove: Use the same cases to compare the current process, the rules-only redesign, and the AI-assisted design. That makes the source of any improvement visible. The 180-case mix is an illustrative learning design, not a formal statistical power calculation. Leadership needs a consistent pattern across routine, complex, and exception cases that is strong enough to support the next decision.



Population and case mix

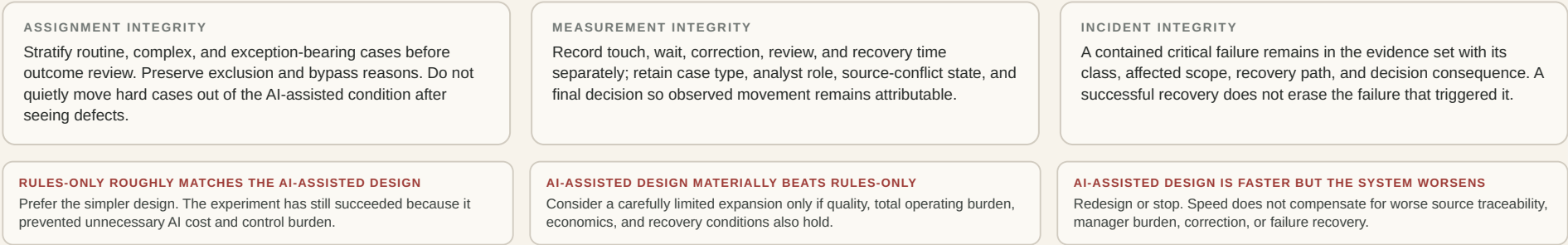
The synthetic design target is **180 eligible cases**, not a formal statistical power calculation: roughly 100 routine, 50 complex multi-field, and 30 exception-bearing cases used live or in controlled non-live stress conditions depending on risk. Results are segmented by class; difficult eligible cases cannot be silently removed after outcomes are visible.

Week 2 retests a fixed same-case set through rules-only and AI-assisted conditions before live reliance. That is the key same-case comparison: it measures the incremental contribution of AI after the rules-based repair.

Entry conditions

All required before live activation: baseline measurement setup; approved source set; case-record schema that preserves the full decision trail; approved tool/data path; versioned change control; named exception and stop owners; proficiency gate; tested manual fallback. Novel policy interpretation, unresolved legal/regulatory applicability, unapproved data paths, and external commitments remain excluded.

Interpretability over theater: the target case count exists to force representative exposure and segmented learning. It is not used to manufacture statistical certainty the design cannot support.



Each week is designed to answer a different leadership question.

The six-week sequence establishes baseline truth, isolates incremental value, exposes live behavior, stresses recovery, and then returns for a decision.

How to use the six weeks: Each week has a different evidence job: establish the baseline, isolate rules-only versus AI-assisted value, introduce limited live cases, expand case complexity, test failure and recovery, and then return to leadership. Advancement depends on evidence, not the calendar. A failed gate means hold, redesign, narrow, or stop.

WEEK 1 - BASELINE AND FREEZE
 freeze pilot scope and eligibility; instrument current-state step timestamps and work minutes; approve source classes/corpus; assign exception/stop owners; complete participant proficiency gate; freeze metric definitions

WEEK 2 - CONTROLLED RETEST SAME-CASE COMPARISON
 retest 60 fixed historical/synthetic cases through current-state representation, rules/forms-only, and AI-assisted design; no operational reliance; compare incremental AI contribution and error classes; repair before live activation

WEEK 3 - LIMITED LIVE START
 activate on routine eligible cases; mandatory named reviewer final decision; retain comparison/holdout; daily defect review

WEEK 4 - EXPAND CASE MIX
 add complex multi-field cases; continue segmentation; test contradiction/handling cases that should be declined or escalated in safe conditions

WEEK 5 - CONTROLLED STRESS AND RECOVERY
 introduce predefined exception-bearing challenge cases in a non-live or otherwise non-harmful setting; exercise stop/contain/retest/resume mechanics; measure manager/SME burden

WEEK 6 - STABILIZE AND DECIDE
 freeze results; reconcile outliers and open incidents; run critical-failure decision gate; return Scale within the proven limits / Redesign and retest / Hold for evidence / STOP

How we keep the comparison fair

Stratify by case type and role; preserve a rules-only or matched comparison; predeclare thresholds; record bypass/exclusion reasons; separate challenge and non-live cases from live outcomes; preserve hard cases; distinguish touch, wait, correction, and verification time. Queue load and analyst mix are documented so cycle-time movement is not automatically attributed to assistance.

What the pilot is not

It is not an adoption campaign, model bake-off, headcount case, or proof that Cedar Ridge is "AI-ready." The pilot is a decision instrument. Its success criterion is whether it returns evidence strong enough to authorize a next step with clear limits or to say stop.

Final return package: evidence-backed return packet identifying what improved, what failed, where burden moved, which controls operated, what remains unknown, and which decision is earned.

GATE BEFORE LIVE USE
After Week 2: Ops + Risk ask whether the rules-only comparison is functioning, critical failures are absent, and incremental AI value remains plausible. If not, live activation does not occur.

GATE DURING LIVE USE
Weeks 3-5: weekly critical-failure review can continue, narrow, contain, or return a case type to role-owned work. Calendar progression never outranks evidence.

GATE AT WEEK 6
Executive return: only SCALE WITHIN PROVEN LIMITS, REDESIGN AND RETEST, HOLD FOR EVIDENCE, or STOP are legitimate. "Pilot completed" is not itself a decision.

RETURN: WHAT IMPROVED
 Show the before/after pattern by case type and identify which gains came from process, rules, and AI.

RETURN: WHAT FAILED
 Surface critical defects, difficult case types, unresolved source conflicts, and any controls that did not operate as designed.

RETURN: WHERE BURDEN MOVED
 Show analyst, manager, SME, correction, escalation, and control effort together with the break-even implication.

RETURN: WHAT LEADERSHIP SHOULD DO
 Recommend scale within proven limits, redesign and retest, hold for evidence, or stop, with the evidence and conditions behind that recommendation.

Speed only counts when quality, evidence, and operating burden remain acceptable.

The scorecard combines completeness, total operating burden, correction, source traceability, and other measures so one attractive metric cannot hide a material failure elsewhere.

How leadership should read success: Read the scorecard as a system. Faster cycle time matters only if completeness, source traceability, correction, and review burden remain acceptable. Segment results by case type so routine wins cannot hide exception failures. The return decision should show both the attractive results and the evidence that argues against expansion.

MEASURE / FORMULA	BASELINE	SCALE ZONE	REVIEW ZONE	STOP ZONE
M01 First-pass completeness eligible cases complete at first analyst review / eligible cases	measure Week 1; synthetic prior signal 66% implied by 34% follow-up, but do not treat as exact equivalent	>=20% relative improvement AND no downstream correction increase	5-19% relative improvement or narrow-case-only benefit	no material improvement plus added burden, or false-complete defects increase
M02 Analyst touch time sum analyst active minutes / correctly resolved eligible cases	Week 1 segmented by case type	>=15% reduction in routine/complex eligible classes	0-14% reduction or highly uneven	increase >10% without offsetting quality/control rationale
M03 Manager/SME burden manager+SME active minutes attributable to pilot / eligible cases	Week 1	no >5% increase overall and no material concentration shift	5-15% increase or short-lived learning burden with declining trend	>15% increase for two review periods or queue harm
M04 Net work minutes per correctly resolved case analyst + manager/SME + correction/rework + incident/recovery minutes / correctly resolved cases	Week 1	>=10% net reduction with quality/control floors met	0-9% reduction	net increase after stabilization
M05 Material decision/classification quality material classification/route errors / reviewed cases, segmented by case type	Week 1 tested manual process	non-inferior to baseline; no new material error class; zero escaped critical decision-control defect	minor contained degradation in known class	critical escaped defect or recurrent material error cluster
M06 Source trace completeness assisted retrievals with source class + owner + version + effective date + link/span / assisted retrievals	new pilot control	>=98% and 100% on relied-upon sources	95-97.9% with contained defects	missing source traceability contributes to material wrong decision or repeated <95%

Critical-failure rule: a green cycle-time or analyst-efficiency metric cannot offset a critical quality, decision-control, source traceability, manager-burden, data-path, or recovery failure. Averages are segmented by case type so a routine-case gain cannot hide exception harm.

JOINT SUCCESS, NOT ONE KPI

Cycle time, quality, source traceability, total operating burden, and correction must move together within acceptable limits.

SEGMENT BY CASE TYPE

Routine-case averages can hide failures in complex or exception work. Review performance by the work that actually drives risk and burden.

LOOK FOR GAMING AND BURDEN SHIFT

A metric that improves because work was deferred, escalated, or pushed to another role should be treated as a warning, not a win.

Scale only if four conditions are true at the same time.

Decision ownership, incremental AI value, conservative economics, and recovery capability are all required before assisted work can expand.

Scale decision rule: Expand only when four conditions hold together: named owners still control consequential decisions; AI adds measurable value beyond the simpler repair; the economics remain positive after the full review and control burden is counted; and the team can recover safely from failure. One failed condition is enough to hold or redesign.

MEASURE / FORMULA	BASELINE	SCALE ZONE	REVIEW ZONE	STOP ZONE
M07 Source-conflict handling correct decisions to decline or escalate / seeded or naturally occurring material source-conflict events	proficiency challenge set	>=95% overall and 100% for critical conflict scenarios	90-94.9%	critical conflict silently resolved by system or systematic under-escalation
M08 Reviewer challenge/override quality appropriate overrides + appropriate decisions to decline or escalate + correct acceptances, adjudicated on review sample / reviewed assisted cases	none	no automation-complacency pattern; unexplained acceptance low and declining	localized complacency corrected by coaching	persistent uncritical acceptance causes material defect
M09 Cycle time elapsed time from preserved receipt to analyst triage decision	Week 1 by case type	>=15% reduction without burden/quality harm	0-14% reduction	worse >10% after stabilization
M10 Incremental AI contribution net burden/quality difference AI-assisted design vs rules/forms-only comparison on matched/retested cases	Week 2 retest	AI-assisted design materially outperforms rules-only on at least one material burden dimension without violating any floor	difference inconclusive	rules-only equal/better after considering verification/governance cost
M11 Economic break-even annualized conservative net capacity value - incremental annual operating/control cost	client assumptions required; public example uses synthetic assumptions only	positive under conservative assumptions after manager burden and correction cost	positive only under mid/high assumptions	negative under plausible conservative case
M12 Recovery proof critical recovery drills completed with evidence / planned recovery drills	0	100% required drills pass including tested manual fallback and retest/resume	noncritical drill defect repaired before return decision	cannot restore safe manual processing or cannot determine affected scope

AI contribution gate: AI-assisted design must materially outperform rules/forms-only on at least one important dimension without violating any floor. "People liked it" and "usage was high" do not satisfy this.

Economic gate: annualized capacity is reported separately from cash savings. A positive capacity estimate requires conservative economics after manager burden, correction, control cost, and plausible attribution limits.

PEOPLE REMAIN IN CONTROL

Consequential decisions, overrides, risk acceptance, and external commitments stay with named owners.

AI ADDS SOMETHING REAL

The AI-assisted design must outperform the rules-only repair on the work it is meant to improve.

THE ECONOMICS SURVIVE FULL BURDEN

Added tools, review, manager time, correction, and controls must still leave a credible net benefit.

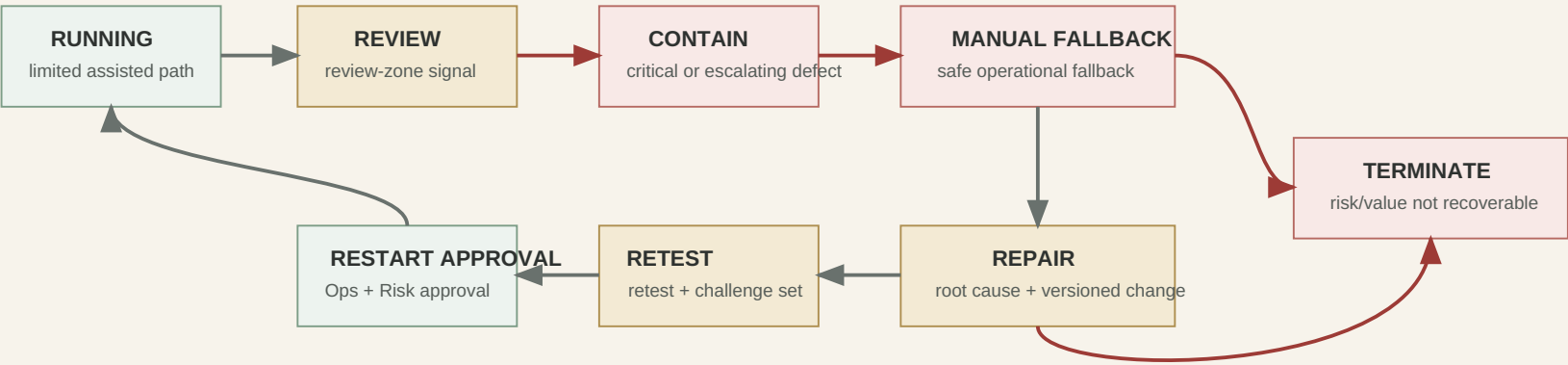
FAILURE IS RECOVERABLE

The team can detect a problem, return to the tested manual process, repair, retest, and resume only through a named gate.

If something goes materially wrong, assisted work stops and the tested manual process takes over.

The recovery path defines detection, containment, tested manual fallback, repair, retesting, and the evidence required before assisted work can resume.

Recovery rule: Define the response before a material failure occurs. Stop assistance for the affected cases, preserve the evidence, return those cases to the tested manual process, identify the failed control or assumption, repair it, retest representative cases, and require Operations and Risk approval before resuming. Recovery is part of the operating design.



Critical stop conditions

Unauthorized external commitment; material wrong decision caused by a missing or stale controlling source; unapproved sensitive-data path; system silently resolves a critical source conflict; tested manual fallback unavailable; repeated critical defect after repair.

Resume conditions

Affected scope known; evidence preserved; root cause or specific corrective hypothesis documented; versioned repair; retest/challenge set passes; source/data/control integrity confirmed; Ops and Risk approve controlled resume; heightened monitoring window defined.

Failure doctrine: a prompt tweak is not a recovery plan. The system contains the affected class, returns work to the safe manual path, preserves evidence, repairs at the failed control or design assumption, retest-tests the repair, then resumes only through a named approval gate.

1. DETECT AND CONTAIN

Identify the affected case type and stop assistance there before the failure spreads.

2. RETURN TO SAFE WORK

Route affected cases to the tested manual process while preserving all evidence and decisions.

3. REPAIR AND RETEST

Fix the earliest failed control or design assumption, then retest representative routine and exception cases.

4. RESUME BY DECISION

Operations and Risk approve resumption only after the repair evidence, source integrity, and fallback readiness are visible.

Even a small pilot changes queues, skills, incentives, precedent, and trust.

Wider effects are monitored so a locally successful pilot does not create a weaker operating system elsewhere.

What to watch beyond the headline metrics: A pilot can improve the measured step while making the wider operation worse. Watch for queues moving elsewhere, skill loss, metric gaming, unofficial workarounds, control fatigue, inconsistent client experience, and savings claims that outrun the evidence. These effects belong in the scale decision because they change the real operating cost and risk.

WIDER EFFECT	MECHANISM	HOW IT BECOMES VISIBLE	DESIGN / RESPONSE
Bottleneck migration	upstream speed increases exception arrival rate to managers	manager queue/touch minutes	narrow case mix, add capacity/ownership, or stop expansion
Deskilling	analysts stop practicing source/exception reasoning	challenge-set decay, shallow overrides, reliance pattern	proficiency refresh, preserve manual-review cases, rotate review
Precedent amplification	retrieval overweights past cases that were context-specific	prior-case citation frequency + conflict review	precedent never controlling; require a linked controlling source and owner review
Metric gaming	teams optimize cycle time/completeness at expense of escalations or quality	manager burden, escaped defects, escalation pattern	scorecard where critical failures cannot be averaged away
Off-system workarounds	staff keep off-system notes because case record is burdensome	decision reconstruction check + missing rationale	reduce capture friction; enforce minimum evidence spine
Control fatigue	too many mandatory review steps make low-risk work slower	verification minutes + bypass reasons	risk-tier verification and remove controls only after evidence
Client experience inconsistency	pilot cases move differently from nonpilot cases	cycle time/complaint/handback by mode	bound communication expectations; no autonomous response
False savings narrative	gross analyst minutes are annualized as cash ROI	finance review of redeployment/cost avoidance	report capacity separately from cash realization

What is genuinely immaterial here: broad enterprise culture transformation is outside this focused Diagnostic. Culture matters only where it changes observable pilot behavior such as challenge, decisions to decline or escalate, bypass, escalation, or manager ownership. That keeps the analysis broad without turning every engagement into an unlimited change program.

LINKED EFFECT · FASTER HANDLING CAN CREATE A QUEUE

Faster analyst handling can raise exception arrivals per hour. If manager capacity is fixed, queue age grows and end-to-end performance can worsen even while local productivity improves.

LINKED EFFECT · TRUST CAN WEAKEN CHALLENGE

As suggestions become reliable, challenge behavior may fall. Lower challenge can hide drift until a rare source conflict appears. Proficiency and recovery drills counter that self-reinforcing pattern.

LINKED EFFECT · MEASURES SHAPE BEHAVIOR

If leaders reward usage or throughput, people optimize those measures. The scorecard therefore protects decisions to decline or escalate, quality, burden, source trace, and recovery as co-equal conditions.

Move to the next stage only when the evidence is strong enough.

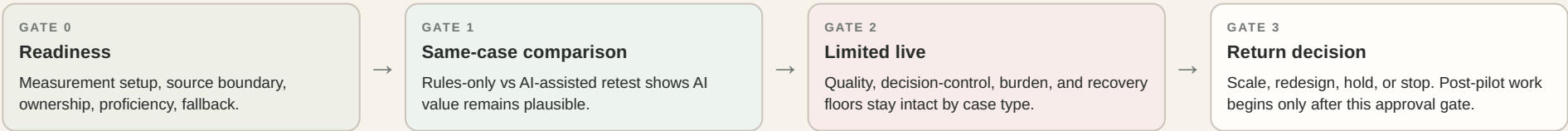
Readiness, same-case comparison, limited live use, stress testing, and scale decisions are organized by evidence, dependencies, and named owners rather than calendar optimism.

What unlocks the next step: Move only when the prerequisite for the next stage is proven: known source boundaries, measured baseline burden, a tested manual fallback, acceptable quality, full-cost economics, and a named decision owner. This lets Cedar Ridge accelerate where the evidence is strong and stop cleanly where it is weak, without treating later phases as preapproved.

HORIZON	WORK	OWNER(S)	DEPENDENCY	GATE	EXIT CONDITION
0-2 weeks	Readiness: intake taxonomy, source registry, case record schema, measurement setup, ownership matrix, approved tool/data path, proficiency challenge set, fallback drill	Ops lead, Risk/Policy Ops, Tool/Data owner	None	Pilot readiness	all entry conditions pass; otherwise HOLD
Week 1	Baseline and freeze	Ops lead	Pilot readiness	Baseline sufficient	segmented baseline and metric definitions complete
Week 2	Rules-only vs AI-assisted controlled retest	Ops lead, Tool/Data owner	Baseline sufficient	Live activation	no critical failure; incremental AI value remains plausible
Weeks 3-4	Limited live pilot routine then complex case mix	Ops lead, Analyst cohort	Live activation	Weekly critical-failure review	continue, contain, or narrow by evidence
Week 5	Controlled stress/recovery proof and manager-burden review	Ops, Risk, Tool/Data owner, Manager/SME	stable live operation	Recovery proof	recovery/fallback pass or no scale recommendation
Week 6	Stabilize analysis and compile return decision	Ops sponsor, Risk, Executive sponsor	complete evidence set	Executive return decision	Scale within the proven limits / Redesign and retest / Hold for evidence / STOP
Post-pilot if scale authorized	Client-owned operating handoff: source/control ownership, metric cadence, change/recheck protocol, incident/recovery, proficiency refresh, backlog for Workflow B prerequisites	Named client operating owners	Executive scale decision	Transfer and sustainment	client can operate/stop/recover/review without hidden Bravura memory

Important dependency: live activation sits after the rules-only vs AI-assisted retest. If AI adds no material incremental value or exposes a critical control defect, Cedar Ridge can stop before exposing the operational queue.

Workflow B stays separate: source-control redesign is a prerequisite backlog, not an automatic phase after Workflow A. A later decision must re-test its own evidence, decision ownership, and economics.



At the end of six weeks, leadership should know enough to choose the next move.

The experiment earns its value by converting today’s uncertainty into a small set of decision-ready facts.

What should come back: a concise evidence package that separates process improvement from incremental AI value, shows where burden moved, documents quality and source performance, tests the economics, and proves whether the operation can fail safely and recover.

1. WHAT DID THE SIMPLER REPAIR ACCOMPLISH?

Evidence: current-state versus rules-only results on the same case mix.
Decision use: establishes how much value came from better intake, routing, and explicit rules before AI receives credit.

2. WHAT DID AI ADD BEYOND THAT?

Evidence: rules-only versus AI-assisted performance on extraction, retrieval, classification support, and total operating burden.
Decision use: shows whether AI contributes enough incremental value to justify its added controls and review work.

3. DID QUALITY AND SOURCE DISCIPLINE HOLD?

Evidence: completeness, source traceability, correction, exception, and source-conflict results by case type.
Decision use: prevents a faster average from hiding a weaker decision process.

4. DID THE ECONOMICS SURVIVE THE FULL BURDEN?

Evidence: analyst, manager, specialist, correction, incident, tool, and control costs against the break-even model.
Decision use: tests whether the apparent efficiency remains worthwhile once hidden work is counted.

5. CAN THE OPERATION FAIL SAFELY?

Evidence: stop, fallback, repair, retest, and restart performance from planned recovery tests and any live incidents.
Decision use: shows whether Cedar Ridge can contain a problem without depending on improvisation.

6. WHAT SHOULD HAPPEN NEXT?

Evidence: the complete pattern, including contrary evidence and unresolved unknowns.
Decision use: supports one explicit recommendation and names what would change that recommendation.

SCALE WITHIN PROVEN LIMITS

Expand only the work that met the quality, value, economics, and recovery conditions.

REDESIGN AND RETEST

Repair a specific weakness, preserve what worked, and rerun the affected comparison.

HOLD FOR EVIDENCE

Do not expand while a decision-critical unknown remains unresolved.

STOP

Use the simpler design or discontinue the AI path when the evidence does not justify it.

Bravura recommendation: do not return with a technology demo. Return with a decision package that shows what improved, what AI added, what control and review work it required, what failed, how recovery performed, and what leadership should do next.

Authorize the experiment, then require evidence before anything expands.

Authorize a limited Workflow A experiment, keep Workflow B closed, and return in six weeks with evidence strong enough to scale, redesign, hold, or stop.

Decision to make now: Authorize the six-week Workflow A experiment and nothing broader. At the end of the six weeks, Bravura should return with evidence on incremental AI value, total operating burden, quality and source traceability, economics, and recovery performance, together with a clear recommendation to scale within proven limits, redesign and retest, hold for more evidence, or stop.

What leadership is authorizing now

Authorize one six-week Workflow A experiment using improved intake and explicit rules as the baseline repair, with AI assistance within defined limits only for eligible unstructured extraction, controlled source retrieval, and classification suggestions.

Require a named decision owner for consequential decisions, preserve source traceability, use the same-case rules-only comparison, count manager and SME burden in the economics, and keep the tested manual process ready as the fallback.

Do not authorize Workflow B interpretation, autonomous external commitments, exception approval, risk acceptance, broad copilot deployment, headcount action, or scale based on adoption or cycle time alone.

DECISION OWNER

Executive sponsor with Operations and Risk. Technical teams may recommend or challenge; consequential decisions remain with named owners.

WHAT SIX WEEKS MUST REVEAL

Whether AI adds material value beyond rules-only; where operating burden moves; how errors differ by case type; whether source-control and decision ownership remains clear and enforced; whether failure is recoverable; and whether the economics survive full burden accounting.

WHAT WOULD CHANGE THIS RECOMMENDATION

Rules-only captures most value; manager burden rises; quality/source traceability deteriorate; source conflicts dominate; recovery fails; or the AI-assisted design cannot demonstrate incremental value.

SCALE WITHIN THE PROVEN LIMITS

Expand only the proven assisted steps and only within the case types, controls, and decision ownership the evidence supports.

REDESIGN + RETEST

Keep the problem open, repair the operating model, and run another limited, reversible experiment where the evidence identifies a fixable failure.

HOLD FOR EVIDENCE

Pause expansion when the decision is still plausible but a missing quantity or unresolved ownership or approval issue could materially change the answer.

STOP

End the assisted path when controls, economics, recovery, or incremental value fail. A stopped pilot is a valid learning outcome.

Return point: leadership receives the same evidence-backed choice after six weeks: scale within the proven limits, redesign and retest, hold for evidence, or stop. Workflow A success does not automatically reopen Workflow B.

Illustrative example only. Every organization detail, metric, threshold, case, cost assumption, and operating scenario is synthetic and is shown to demonstrate reasoning quality, not client results.

WHY THIS PATH

It creates useful evidence quickly without requiring Cedar Ridge to bet on broad adoption.

WHY NOT MORE

The current evidence does not justify broader decision delegation or Workflow B interpretation.

WHAT GOOD LOOKS LIKE

A six-week return decision with enough evidence to choose scale, redesign, hold, or stop without inventing certainty.