MatchPoint Research Working Paper No. 2026/T18

Agentic Workflows for the Back Office: Reconciliation, Reporting and Compliance

An evidence-led operating and control framework for GCC family businesses and family offices

Prepared by Matchpoint Partners
Author attribution pending CK approval
August 2026

ABSTRACT

Background. Agentic systems can interpret unstructured evidence and invoke tools across multi-step back-office work, while their reliability, authority and security constraints remain task-specific. **Objective.** This paper develops an evidence-led framework for reconciliation, reporting and compliance workflows used by GCC SME and family-business owners and family-office investment teams. **Approach.** The analysis reviews 40 primary, regulatory, standards, research and provider sources available through 1 August 2026 and defines the accepted back-office work packet as the operating unit. **Findings.** Productive deployment requires controlled source lineage, deterministic calculations, least-privilege tools, visible exceptions, representative evaluation, human acceptance and tested recovery. **Implications.** Agents can interpret, assemble and route; systems of record and named professionals retain calculation, posting, filing, payment and compliance authority. Attributed Matchpoint or client revenue, cost reduction and loss reduction remain USD 0 until approved observed attribution exists.

Highlights

Unit. An accepted back-office work packet binds source records, calculations, exceptions, review and release.

Authority. Agents interpret, assemble and route; named owners accept and controlled systems record.

Evaluation. Completeness, authority and material-exception recall precede narrative quality and speed.

Security. Purpose-bound identities, narrow tools, typed outputs, evidence logs and tested recovery constrain agency.

Economics. Value is measured per accepted packet with review, rework, control and failure effort included.

JEL Classification: C88, D24, L86, M15, M41, M42, O33

Keywords: agentic workflows, reconciliation, management reporting, compliance, internal controls, audit evidence, family office, GCC family business, AI governance, process automation, eInvoicing

Research paper only. This document is not investment, legal, regulatory, accounting, audit, tax, privacy, cybersecurity, valuation or technology advice. All worked inputs, thresholds, times, rates and economics are unverified illustrative management assumptions.

1. INTRODUCTION

Back-office work determines whether a business can close its books, understand cash, explain performance, meet tax and regulatory obligations, and give decision-makers information they can trust. Reconciliation, reporting and compliance are therefore authority-bearing processes. Errors can misstate cash, duplicate a payment, conceal an exception, omit a filing input, expose personal data or weaken the evidence available to management, auditors, lenders and regulators.

Agentic systems create a new way to organise this work. An agent combines a model, tools and instructions, and can use those tools over multiple steps to pursue a defined task [5]. This capability can interpret unstructured material, gather evidence from several systems, perform controlled calculations, draft a report and route exceptions. It also expands the operational attack surface. Prompt injection, excessive agency, insecure output handling, credential misuse and compromised tools can turn an apparently useful workflow into an unauthorised actor [10-13].

The commercial question is consequently precise: which back-office work can be converted into evidence-backed, accepted work packets without weakening record integrity, professional judgement or delegated authority? The unit of value in this paper is the **accepted back-office work packet**. A packet contains a defined task, source records, transformations, deterministic calculations, agent trace, exception disposition, reviewer evidence and release status. A draft becomes an accepted packet only after required validations and approvals pass.

This paper is written for B4 GCC SME and family-business owners and A2 family-office CIOs and heads of alternatives. B4 operators need practical gains in close speed, working-capital visibility, tax readiness and management capacity. A2 leaders need consolidated reporting, look-through evidence, capital-call and distribution control, service-provider oversight and an audit trail across entities, portfolios and jurisdictions. Their systems and professional obligations differ. The control logic is shared.

The public evidence supports disciplined experimentation. Stanford HAI reports that organisational AI adoption reached 88% in 2025, while real-computer agents still failed roughly one in three tasks on the cited OSWorld benchmark [1]. A field study of 5,179 customer-support agents found a 14% average productivity increase from a generative-AI assistant, with substantial variation by worker experience [2]. A separate randomised study of experienced open-source developers found that early-2025 AI tools increased completion time by 19% in its setting [3], and the researchers later reported that selection effects made a follow-on estimate unreliable [4]. These results do not measure reconciliation, reporting or compliance. They show why a local baseline, representative cases and observed acceptance are required.

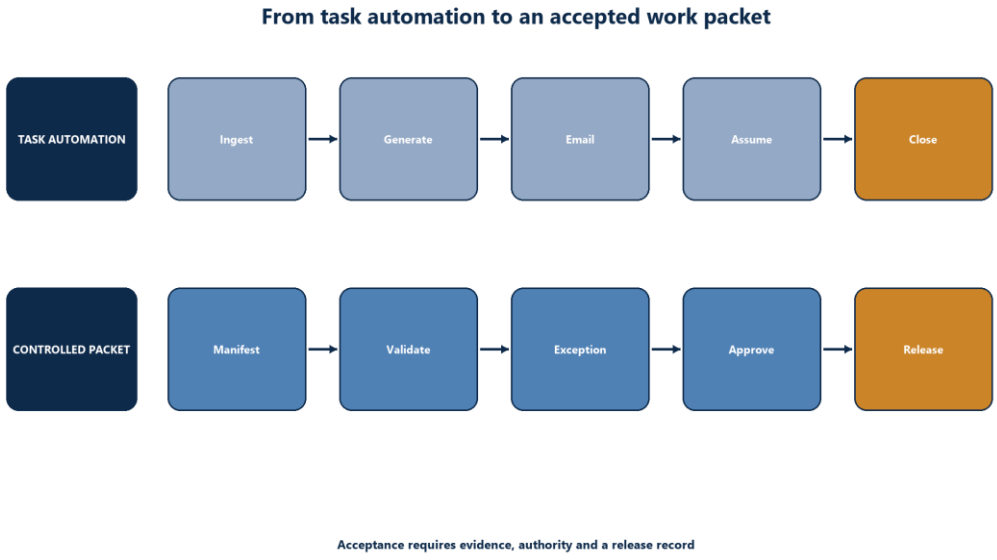

Figure 1. Before and after: task automation versus an evidence-backed work-packet workflow
Source: Matchpoint operating framework.

The framework has six conclusions. First, deterministic controls retain authority over calculations, permissions, posting and release. Second, an agent can interpret, assemble and route; it should not become the system of record. Third, every material statement needs source lineage and a reproducible transformation. Fourth, initial deployment should operate in shadow or draft mode with narrow tools and human approval. Fifth, productivity is measured per accepted packet with comparable quality and control effort. Sixth, attributed Matchpoint or client revenue, cost reduction and loss reduction remain USD 0 until approved observed attribution exists.

Decision	Evidence required	Initial authority
Use an agent for a workflow	Stable task boundary, representative cases and named owner	Process owner
Connect a source system	Purpose, data rights, least privilege and test evidence	System and data owners
Accept a reconciliation packet	Complete population, deterministic tie-out and resolved exceptions	Finance controller
Release a management report	Source lineage, approved definitions and reviewer sign-off	Reporting owner
Clear a compliance case	Policy, evidence, escalation and qualified judgement	Compliance owner
Post a journal, pay or file	Delegated authority and controlled source-system action	Authorised human or separately approved deterministic control

2. SCOPE, DEFINITIONS AND EVIDENCE BOUNDARIES

2.1 Agentic workflow

An agentic workflow uses a model to choose or sequence actions within an approved boundary. A deterministic workflow follows pre-defined code paths. Anthropic distinguishes workflows, in which models and tools follow pre-defined orchestration, from agents, in which the model directs process and tool use [6]. Both can be useful. Reconciliation logic, tax calculations, access checks and posting rules usually benefit from deterministic implementation. Document interpretation, evidence collection, exception description and reviewer routing can benefit from bounded model judgement.

System class	Execution logic	Suitable back-office role	Principal control
Deterministic automation	Rules and code define every step	Matching, calculations, validations and scheduled extracts	Versioned rules and tests
Copilot	User asks and reviews each output	Analysis, narrative drafting and policy retrieval	Visible sources and user acceptance
Tool-using agent	Model selects approved tools within a task	Evidence gathering and exception packet assembly	Tool allowlist, scopes and stop conditions
Orchestrated workflow	Code controls stages; models perform bounded steps	Close, reporting and compliance packet production	Stage gates and independent validators
Autonomous actor	Model can initiate and release material actions	Outside initial scope	Separate authority decision and enhanced control case

2.2 Accepted back-office work packet

The packet is an evidence object and an operational hand-off. It has an immutable identity, task contract, entity and period, source manifest, input hashes or provider identifiers, rules and model versions, calculations, generated narrative, exceptions, approvals, release target and retention status. The evidence ledger records what occurred. The enterprise-resource-planning, portfolio-accounting, banking, tax or compliance system remains the authoritative record.

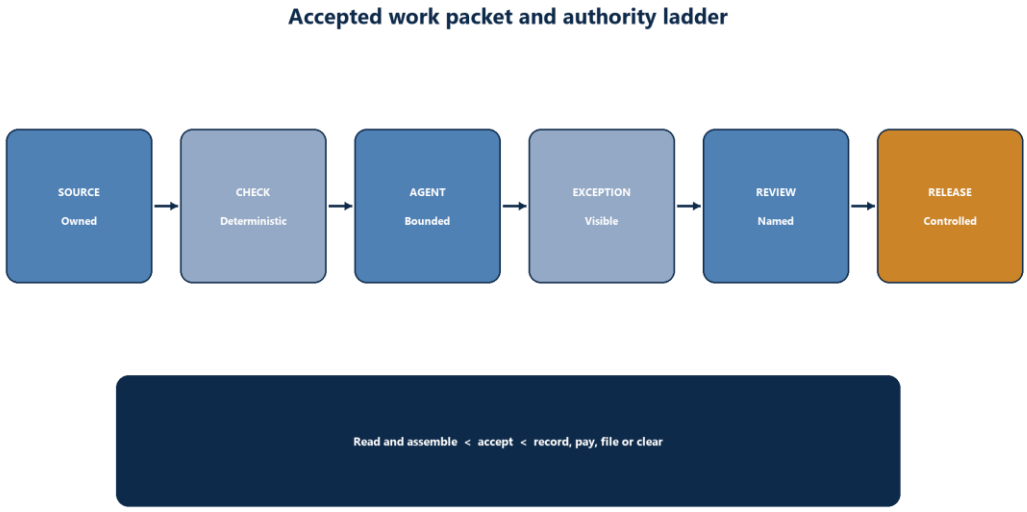

Figure 2. Accepted work packet and authority ladder

Source: Matchpoint control framework.

Acceptance is role-specific. A preparer can confirm completeness and explain an exception. A controller can accept a reconciliation. A portfolio-reporting owner can approve a performance pack. A compliance officer can clear or escalate a case. External auditors and regulators make their own decisions. Agent output cannot pre-approve their conclusions.

Packet state	Meaning	Permitted action
Created	Task contract and source window fixed	Gather evidence
Assembled	Required sources present or gaps recorded	Run deterministic checks
Evaluated	Quality, control and security tests completed	Route exceptions
Reviewed	Named reviewer records disposition	Prepare release
Accepted	All required gates pass	Release the approved packet
Rejected	Material defect or authority failure	Correct, rerun or revert
Superseded	Later approved packet replaces it	Retain lineage; prevent reuse

2.3 Workflow scope

The framework covers record-to-report reconciliations, bank and cash reconciliation, intercompany matching, receivables and payables control, consolidation support, management and portfolio reporting, covenant monitoring, KYC and sanctions evidence collection, tax-record readiness, policy surveillance and compliance case assembly. It excludes autonomous legal interpretation, accounting-policy selection, investment decisions, suspicious-activity filing decisions, sanctions clearance, tax filing, journal posting, payments and changes to master data during initial deployment.

The exclusion does not imply that technology cannot support those activities. It establishes the authority boundary for this paper. A separate design, legal and professional review is required before any organisation changes that boundary.

2.4 Evidence classes

Public sources support framework design and external requirements. Internal operating evidence establishes whether a particular implementation works. Management scenarios illustrate arithmetic only. Each material claim receives an evidence class.

Class	Description	Permitted use
P1	Law, regulation or binding rule	Define an applicable obligation after qualified scope review
P2	Official standard, regulator guidance or authoritative framework	Design controls and review questions
P3	Peer-reviewed or primary empirical research	Bound a claim to the studied task and population
P4	Provider documentation or technical specification	Describe a versioned capability or interface
I1	Reconciled internal source-system data	Measure local operations after ownership and quality checks
I2	Approved test, trace, exception and acceptance record	Evaluate local workflow performance
U	Unverified illustrative management assumption	Demonstrate a formula; never represent observed value

Public evidence cannot prove a local business result. A provider description cannot establish a control's operating effectiveness. A successful demonstration cannot establish representative accuracy or audit sufficiency. An approved internal packet can support a local conclusion only for its defined task, data, period, controls and reviewer.

2.5 Professional and author boundaries

The paper provides a research and operating framework. It does not provide legal, regulatory, accounting, audit, tax, privacy, cybersecurity, investment or valuation advice. Qualified professionals determine applicable obligations and exercise reserved judgement. Named-person authorship remains pending CK approval. All worked values, thresholds, acceptance rates and economics are unverified illustrative management assumptions.

3. B4 AND A2 DECISION MAP

3.1 B4 GCC SME and family-business decisions

B4 operators frequently manage entity complexity with constrained finance, tax, compliance and technology teams. The operating priority is a reliable close and cash view that supports collections, purchasing, borrowing, investment and shareholder decisions. A family business can also require related-party, shareholder-current-account and intercompany evidence that survives succession, financing and transaction diligence.

B4 decision	Packet	Minimum evidence	Authority
Which cash differences require action?	Bank reconciliation	Bank source, ledger source, timing rule, exception owner	Finance controller
Which receivables need collection or dispute handling?	Receivables exception	Invoice, delivery, credit note, payment and correspondence links	AR owner
Can the monthly close be released?	Close control	Reconciliation completion, journals, open exceptions and sign-offs	CFO/controller
Is tax documentation complete?	Tax readiness	Transaction record, invoice, classification, calculation and retention status	Tax owner/adviser
Is an intercompany balance accepted?	Intercompany packet	Counterparty confirmation, currency, period, eliminations and dispute	Entity controllers
Is a lender report ready?	Covenant packet	Facility definition, source balances, calculation and variance explanation	CFO and facility owner

The first B4 deployment should target one repetitive, high-volume task with clear source systems and a named reviewer. Bank reconciliation and monthly management-report assembly are common candidates. A process with unresolved master-data ownership, unstable accounting policy or unclear delegation requires remediation before agentic orchestration.

3.2 A2 family-office CIO and alternatives decisions

A2 teams aggregate information from administrators, custodians, banks, general partners, operating companies and internal books. Reporting dates, valuation conventions, currencies and entity structures differ. The required output may include net asset value, exposure, cash forecasts, capital calls, distributions, unfunded commitments, fees,

liquidity and performance. Agentic assembly can reduce manual document handling. Portfolio accounting, valuation approval and investment judgement retain their established authorities.

A2 decision	Packet	Minimum evidence	Authority
Is the portfolio report complete?	Reporting packet	Entity map, custodian/admin files, valuation dates, FX and reconciliations	Reporting owner
Is a capital call valid and funded?	Capital-call packet	Notice, fund terms, bank instructions, approval and cash forecast	Investment ops and signatories
Are fees consistent with terms?	Fee-review packet	LPA/IMA term, calculation base, period and administrator record	Fund controller/CIO delegate
Does look-through exposure meet policy?	Exposure packet	Holdings lineage, classification, denominator and limit logic	Risk/CIO
Does a manager report require escalation?	Exception packet	Missing data, stale valuation, covenant, key-person or liquidity evidence	Relationship owner/CIO
Can a board pack be released?	Governance packet	Approved metrics, narrative sources, conflicts and sign-offs	CIO/board-secretariat owner

3.3 Shared control questions

Both ICPs should answer the same eight questions before production use: What exact decision is supported? Which records are authoritative? What can the agent read? What can it write? Which calculation is deterministic? Which exceptions stop the workflow? Who accepts the packet? How can the process revert safely?

Question	Weak answer	Gate-ready answer
Objective	“Automate finance”	“Prepare draft daily bank-reconciliation packets for accounts X-Y by 10:00”
Population	“All transactions”	Source, entity, account, date and completeness control defined
Authority	“Finance approves”	Named role, amount/exception thresholds and evidence captured
Data rights	“Already in the system”	Purpose, access, retention and transfer approved by owners
Quality	“The output looks right”	Gold set, independent checks and acceptance threshold fixed
Security	“Vendor is enterprise-grade”	Threat model, scopes, logging, incident and exit tested
Value	“Hours saved”	Comparable accepted packets and full human/system effort measured
Recovery	“We can do it manually”	Tested fallback, last safe state and recovery owner recorded

4. MARKET, ADOPTION AND PRODUCTIVITY EVIDENCE

4.1 Adoption and capability

Stanford HAI reports 88% organisational AI adoption in 2025 and rapid improvement on several technical benchmarks [1]. The same report describes a jagged capability frontier: systems can perform strongly on difficult benchmarks while failing apparently simple tasks, and the cited OSWorld agent benchmark still recorded failure in roughly one third of attempts [1]. A back-office design should therefore treat general capability as context and local evaluation as authority.

The UAE has a strong adoption context. Stanford HAI reports population-level generative-AI adoption of 54% in the UAE and workplace use exceeding 80% in several countries including the UAE and Saudi Arabia [1]. These statistics describe reported adoption. They do not establish production control maturity, economic return or suitability for a named workflow.

4.2 Productivity is task-specific

Brynjolfsson, Li and Raymond found a 14% average productivity increase for customer-support agents using an AI assistant, with larger gains among novice and lower-skilled workers and minimal effects among highly skilled

workers [2]. The treatment provided conversational guidance in a defined setting. It did not autonomously reconcile ledgers or release compliance reports.

METR's randomised study found a 19% slowdown when experienced open-source developers used early-2025 AI tools for real issues in repositories they knew well [3]. The result was explicitly bounded to that setting. A 2026 update explained that participation selection made a later estimate unreliable [4]. Together, the studies show that user perception, benchmark performance and observed elapsed time can diverge.

Evidence	Observed setting	Result	T18 implication
Stanford AI Index [1]	Cross-economy adoption and benchmarks	High adoption; uneven agent reliability	Test local work; preserve exception handling
NBER W31161 [2]	Customer support, 5,179 agents	14% average productivity increase	Assistance can transfer practice in a defined task
METR 2025 [3]	Experienced developers, 246 issues	19% longer with AI	Measure elapsed time and rework directly
METR 2026 update [4]	Follow-on developer experiment	Selection impaired estimate	Record participation and workflow-selection effects

4.3 Productivity measurement contract

The measurement unit is an accepted packet, not a generated output, prompt, agent run or document. The baseline and assisted process should cover the same representative population, quality bar, control obligations and deadline. Total assisted effort includes data preparation, failed runs, reviewer time, corrections, exception handling, control operation, vendor management and recovery.

Let baseline hours per accepted packet be **Hb**. Let assisted preparation, review, exception and control hours be **Ha**. Let accepted packets be **Q**. Gross hours released equal **Q x (Hb - Ha)**. Observed labour-cost reduction requires an approved change in paid cost. Capacity redeployment is reported separately.

Metric	Definition	Exclusion risk
First-pass acceptance	Packets accepted without correction / reviewed packets	Hiding rejected or abandoned packets
End-to-end cycle time	Trigger to accepted release	Reporting model latency only
Human touch time	Preparation, review, correction and exception minutes	Excluding supervision and support
Rework rate	Corrected packets / generated packets	Counting only successful runs
Missed-exception rate	Known material exceptions not raised / known exceptions	Using an incomplete gold set
False-positive rate	Raised exceptions judged immaterial / raised exceptions	Changing reviewer threshold mid-test
Control effort	Hours spent on access, logs, tests, incidents and changes	Treating governance as sunk overhead
Failure severity	Maximum approved loss/impact category of a failure	Averaging away tail risk

4.4 Process selection

Agentic work suits tasks containing unstructured evidence, multiple systems, conditional routing and a clear completion state [5,6]. A deterministic workflow can be more appropriate when inputs, rules and outputs are stable. Process selection should score value, ambiguity, data readiness, reversibility, authority and failure impact.

Factor	Low-risk candidate	High-risk candidate
Input	Structured, complete, owned	Missing, conflicting or unowned
Rule	Explicit and versioned	Professional judgement or unresolved policy
Action	Draft or reversible	Payment, filing, posting or destructive write
Exception	Known classes with owner	Novel, open-ended or time-critical
Review	Named reviewer with capacity	Diffuse ownership or rubber-stamp risk

Factor	Low-risk candidate	High-risk candidate
Impact	Limited and recoverable	Legal, financial, customer or systemic consequence

5. OPERATING MODEL: THE ACCEPTED WORK PACKET

5.1 Task contract

Every workflow begins with a machine-readable task contract. It states the business objective, entity, account or portfolio, period, population, sources, rules, thresholds, tools, permissions, expected packet, stop conditions, reviewer, deadline, retention and recovery process. A natural-language prompt can explain the task. It should not be the only control specification.

Contract field	Example for bank reconciliation	Control purpose
Objective	Draft daily reconciliation packet	Prevent task expansion
Population	Accounts 101-104; previous business day	Establish completeness
Sources	Bank API snapshot and ERP ledger extract	Identify authoritative records
Match rules	Exact reference/amount; approved date window	Make calculations reproducible
Exceptions	Duplicate, missing, stale, currency or threshold	Route known failure classes
Tool rights	Read sources; write packet store only	Limit agency
Stop	Missing source, failed hash, imbalance over threshold	Fail closed
Acceptance	Controller approval with exception disposition	Preserve authority

5.2 Evidence manifest

The manifest records each source object, provider identifier, extraction timestamp, event timestamp, entity, period, schema, checksum or immutable version, access path, quality result and retention class. It distinguishes external records from company-produced information. PCAOB AS 1105 requires auditors using company-produced information to evaluate accuracy, completeness and precision, and addresses the reliability of external electronic information provided by the company [17]. The standard applies to PCAOB audits. Its evidence logic is useful as a design question outside that jurisdiction; it does not make an internal packet audit evidence automatically.

Manifest block	Required fields	Reviewer question
Identity	Source, object, owner and system	What is this record?
Scope	Entity, period, account, portfolio and population	Is the population complete?
Provenance	Created, extracted, received and transformed times	Where did it originate?
Integrity	Hash, provider ID, version and immutability state	Has it changed?
Rights	Purpose, role, legal basis if applicable and retention	May it be used here?
Quality	Schema, completeness, reconciliation and exception	Is it fit for this task?

5.3 Deterministic calculation layer

Arithmetic, matching, eligibility, threshold and policy rules should execute in versioned code or rules engines where feasible. The agent can propose an interpretation, select an approved rule or explain a result. Independent validation recomputes material values from the accepted source snapshot. The packet retains both raw and calculated fields.

For a cash reconciliation, the closing equation is opening ledger cash plus ledger movements equals closing ledger cash. The bank side has its own opening, movements and closing balance. The reconciliation explains timing and permanent differences. The agent may classify a narrative reference. It may not alter the population or force a tie.

5.4 Exception queue

An exception is a first-class record with identity, source links, category, value, age, severity, proposed action, owner, due date, status and reviewer disposition. A workflow that produces a polished narrative while suppressing unresolved exceptions has failed.

Exception class	Example	Required disposition
Completeness	Missing bank statement or manager file	Stop or approve explicit limitation
Integrity	Hash or sequence mismatch	Quarantine and investigate
Rule	No approved classification or accounting treatment	Escalate to policy owner
Threshold	Amount, age or concentration exceeds limit	Named authority decision
Contradiction	Source records disagree	Resolve or report both with limitation
Security	Unexpected tool, prompt injection or credential event	Stop, preserve trace and invoke incident process
Reliability	Validator or model below threshold	Revert to approved process

5.5 Acceptance record

Acceptance records the reviewer, role, timestamp, packet version, exceptions, approvals, limitations and release target. The interface should present decisive evidence, not a long agent transcript. The full trace remains available to investigation and assurance teams under access and retention controls.

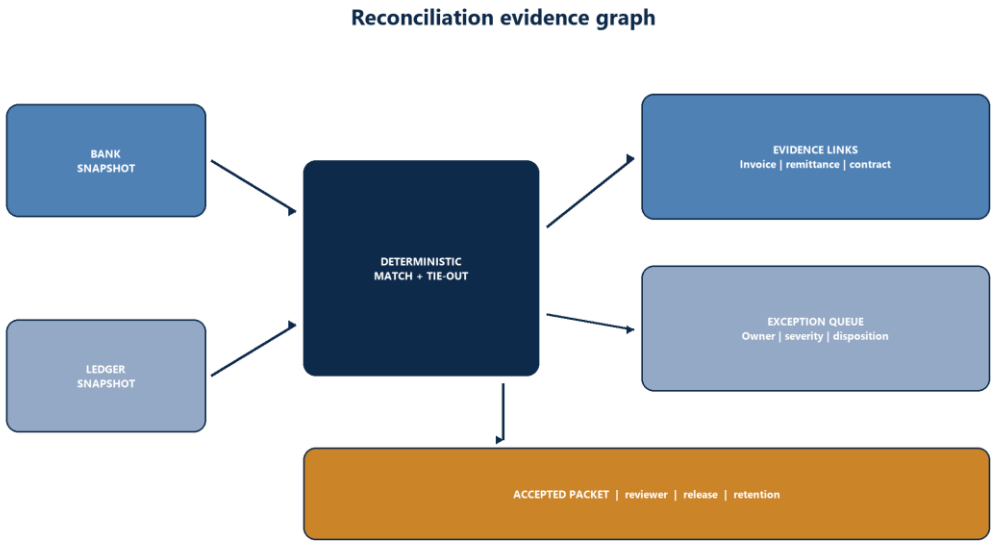

Figure 3. Reconciliation evidence graph from source record to accepted packet

Source: Matchpoint evidence architecture; audit-evidence context [15-18].

6. THREE REFERENCE WORKFLOWS

6.1 Reconciliation workflow

The reconciliation workflow gathers a frozen source population, validates completeness, normalises identifiers, applies deterministic match rules, proposes classifications for residuals, links supporting evidence, routes exceptions and prepares an acceptance packet. The agent can search an approved document repository for remittance, invoice or contract evidence. It cannot fabricate a missing document or post a correcting journal.

Stage	Deterministic component	Agentic component	Human authority
Ingest	Count, totals, sequence and schema	Explain missing or anomalous inputs	Source owner resolves gaps

Stage	Deterministic component	Agentic component	Human authority
Match	Exact and approved fuzzy rules	Suggest narrative/category	Controller approves material residuals
Evidence	Hash and link validation	Retrieve approved supporting items	Preparer confirms relevance
Exception	Threshold and ageing	Draft cause and action	Named owner disposes
Release	Recompute and completeness gate	Draft summary	Controller accepts
Post-close	Lock packet and monitor	Draft recurring-pattern insight	Process owner changes rules

Duplicate detection deserves special treatment. Similar amount, date and counterparty can indicate a duplicate or a legitimate recurrence. The packet should show the features, linked source documents and rule outcome. Payment cancellation, recovery or posting remains under authorised workflow.

6.2 Reporting workflow

The reporting workflow defines a metric dictionary before generation. Each metric has name, calculation, source, owner, frequency, dimension, unit, tolerance and presentation rule. Narrative statements link to metric records and source evidence. Structured reporting reduces ambiguity: the IFRS Foundation describes digital financial reports as computer-readable structured data and maintains taxonomies that support implementation and use [40]. The taxonomy does not determine internal management definitions.

Reporting layer	Controlled object	Failure prevented
Entity map	Legal and reporting entities, ownership and consolidation role	Omission or double counting
Metric dictionary	Formula, source, unit and owner	Definition drift
Period calendar	Close dates, valuation dates and cut-offs	Stale or mixed-period reporting
FX table	Source, timestamp and translation rule	Inconsistent conversion
Narrative claim	Metric links, comparison and limitation	Unsupported commentary
Release pack	Version, reviewers, audience and distribution	Wrong or premature disclosure

An A2 portfolio report should distinguish reported manager values, administrator values, custodian positions and internally approved adjustments. Stale values and estimated look-through data are labelled. An agent can extract and compare documents. Valuation acceptance remains with the appointed authority.

6.3 Compliance workflow

The compliance workflow assembles evidence, applies deterministic screening or rules, identifies discrepancies, drafts a case summary and routes it. FATF states that new technologies can improve AML/CFT speed, quality and efficiency when implemented responsibly and through a risk-based approach, with privacy, data protection, informed oversight and cooperation [37]. FFIEC guidance on automated suspicious-activity monitoring emphasises defined filtering criteria, controlled changes, periodic testing and independent validation [39]. OFAC's framework identifies management commitment, risk assessment, internal controls, testing/auditing and training as core sanctions-program components [38].

Compliance activity	Agent support	Reserved decision
KYC refresh	Extract, compare and request missing evidence	Customer risk acceptance
Sanctions alert	Assemble names, identifiers, geography and source records	True-match/clearance decision
Transaction monitoring	Explain rule outputs and related activity	Suspicion and filing decision
Policy surveillance	Identify source change and affected procedures	Legal interpretation and policy approval
Tax readiness	Assemble transactions, invoices and retention evidence	Tax treatment, return and filing
Privacy request	Locate controlled data and route work	Legal scope, disclosure and redaction

The agent should not receive the power to suppress an alert, file a report, release blocked funds or make a significant personal-data decision during initial deployment. It should preserve all competing evidence and record uncertainty.

7. ARCHITECTURE AND TOOL STACK

7.1 Control-plane architecture

The architecture separates systems of record, data movement, policy, agent reasoning, deterministic validation, evidence storage, review and release. The separation reduces the risk that a fluent model output becomes an operational record without independent checks. It also creates a location for model or provider substitution.

Figure 4. Agentic back-office control-plane architecture and tool stack

Source: Matchpoint architecture; agent, security and resilience context [5-13,21-25,31].

Layer	Components	Required evidence
Systems of record	ERP, banks, portfolio accounting, CRM, document and compliance systems	Owner, interface, schema and availability
Ingestion	API, event stream, approved file exchange and document capture	Completeness, sequence, time and integrity
Identity and policy	Users, service identities, roles, matters, entities and purposes	Least privilege and segregation of duties
Deterministic services	Matching, calculations, limits, schema and policy rules	Code/rule version, tests and owner
Agent control plane	Model routing, tools, instructions, memory and stop conditions	Version, scope, trace and evaluation
Evidence ledger	Manifests, transformations, outputs, exceptions and approvals	Immutability, retention and access
Review and release	Queue, dual control, source-system adapter and notification	Named authority, idempotency and rollback
Observability	Runs, tool calls, latency, cost, errors, drift and incidents	Alerts, thresholds and response owner

7.2 Identity, rights and segregation

Each agent run should use a purpose-bound service identity with only the tools and records required for the named task. The MCP specification requires scope-aware authorisation, token audience validation and secure token handling, and prohibits token passthrough in the cited current specification [13]. The underlying principle applies across integration protocols: a credential issued for one resource should not become a universal key.

Segregation of duties applies to humans and services. A workflow that prepares a journal cannot approve or post it. A workflow that assembles payment evidence cannot change beneficiary master data or release the payment. A compliance case assembler cannot change screening rules or clear its own alerts.

Capability	Preparation identity	Approval identity	Release identity
Reconciliation	Read sources; write packet	Read packet; approve/reject	Controlled adapter after acceptance
Reporting	Read controlled metrics; draft narrative	Approve definitions and pack	Publish to named audience
Compliance	Read case sources; assemble	Clear/escalate by delegated role	Separate filing/blocking channel
Configuration	Propose a rule or prompt change	Change owner and validator approve	Deployment pipeline releases

7.3 Tool contracts

A tool contract describes its business purpose, exact inputs and outputs, authentication, scopes, side effects, idempotency, timeouts, retries, rate limits, error classes, logging, test environment and owner. Tool descriptions are untrusted inputs unless obtained from a trusted source [13]. The orchestration layer should validate parameters against schemas and reject unexpected tools or changed capabilities.

Read tools and write tools should be separated. Early workflow stages use read-only connections. Draft outputs write to a controlled packet store. Source-system writes occur only through narrow, approved release adapters after acceptance. Destructive or irreversible actions stay disabled unless a separately approved use case requires them.

7.4 Evidence ledger

The evidence ledger is append-only at the packet level. Corrections produce a new version linked to the superseded version. The ledger stores input identities, transformations, deterministic outputs, model/provider identifiers, instructions, tool calls, validation results, exceptions, approvals and releases. Sensitive content can be stored by reference with access controls; the trace should not become an uncontrolled duplicate data lake.

Ledger event	Minimum record	Integrity control
Task created	Contract hash, owner and deadline	Signed/versioned contract
Source received	Provider/object ID, timestamp and hash	Completeness and source validation
Tool invoked	Identity, tool, parameters, result and duration	Schema, allowlist and scope check
Model executed	Provider, model, instruction and output reference	Version and trace ID
Validator executed	Rule/test version and result	Independent implementation where material
Exception changed	State, owner, evidence and reason	Role and chronology
Packet accepted	Reviewer, role, limits and disposition	Authentication and dual control where required
Released	Target, adapter, idempotency key and result	Reconciliation back to target

7.5 Reliability and recovery

The workflow must fail safely when a source, provider, tool, validator or reviewer is unavailable. The Basel operational-resilience principles define operational resilience around delivery of critical operations during disruption and emphasise protection, detection, response, recovery and testing [21]. CBUAE operational-risk standards require identification and assessment of risk across material products, activities, processes and systems [22]. For a bank or regulated financial institution, applicable mandatory requirements require qualified mapping. For an SME or family office, the principles remain useful design inputs.

Recovery maintains the last safe state, incomplete packet, exception ownership and manual or deterministic fallback. Queue backlogs receive limits. A provider outage cannot silently convert a daily controlled process into an unreviewed batch later.

8. EVALUATION, VALIDATION AND RELEASE

8.1 Evaluation contract

The evaluation contract fixes the task family, population, gold set, holdout set, adverse cases, metrics, thresholds, severity scale, reviewers, statistical treatment and change rule before results are known. NIST's AI RMF and

Generative AI Profile organise voluntary risk work across govern, map, measure and manage functions [7,8]. The IIA's AI Auditing Framework describes governance, management and internal-audit considerations for AI [16]. These frameworks inform the programme; they do not certify a named workflow.

Evaluation set	Purpose	Example
Gold	Establish accepted answer and evidence	Reconciliations independently completed by qualified staff
Holdout	Test generalisation without tuning leakage	Later periods, entities and counterparties
Adverse	Test manipulation and failure	Prompt injection in invoice text; conflicting bank record
Boundary	Test authority and policy	Request to post, pay, clear or file
Recovery	Test outage and partial completion	Model unavailable after source capture
Drift	Test changed formats and operating patterns	New administrator template or eInvoice field

8.2 Metric hierarchy

Completeness and authority come before prose quality. A packet fails if it omits a source population, changes a controlled calculation, hides an exception, exceeds tool rights or lacks required approval. Narrative accuracy is then tested at claim level. Operational metrics follow.

Priority	Metric	Gate example
1	Population completeness	100% of required source objects or explicit stop
2	Deterministic tie-out	Exact equality within approved currency precision
3	Authority adherence	Zero unauthorised writes or releases
4	Material exception recall	Threshold set by risk owner; misses severity-weighted
5	Claim support	Every material statement links to evidence
6	First-pass acceptance	Measured against frozen representative set
7	Cycle and touch time	End-to-end and human effort both recorded
8	Cost and resilience	Full run, review, support and failure cost

Threshold values depend on the workflow. Any threshold shown in a worked example is an unverified illustrative management assumption until the named risk owner approves it.

8.3 Shadow mode

Shadow mode processes live or representative inputs without changing operational records. Staff perform the approved baseline in parallel. Reviewers compare completeness, exceptions, quality, time and effort. The programme retains disagreements, failed runs and abandoned packets. Selective reporting of successful cases invalidates the evaluation.

Shadow result	Action
Both agree; accepted	Record agreement and effort
Agent finds valid exception missed by baseline	Investigate baseline control and gold set
Baseline finds exception missed by agent	Severity review; block release; update test set
Different classification; same financial result	Policy owner resolves and documents rule
Agent produces unsupported narrative	Reject claim; strengthen evidence binding
Agent attempts unauthorised action	Security incident; suspend workflow and preserve trace

8.4 Change validation

Model, provider, prompt, tool, schema, rule, policy and source changes can alter outcomes. Each material change triggers regression testing against frozen cases and relevant adverse cases. Provider release notes are inputs, not validation. Emergency changes receive time-bound approval and retrospective review.

The IAASB's July 2026 project update states that exposure drafts for proposed revisions to ISA 330, ISA 500 and ISA 520 were approved and were due for public consultation; the proposals address audit evidence, controls, professional scepticism and technology [18]. These are proposals, not final current standards. An implementation should obtain current professional advice for its audit context.

8.5 Release gate

A restricted production release requires the task contract, source controls, tool permissions, evaluation threshold, shadow results, reviewer capacity, incident process, fallback and change process to pass. Release remains limited by entity, period, account, portfolio, value and action.

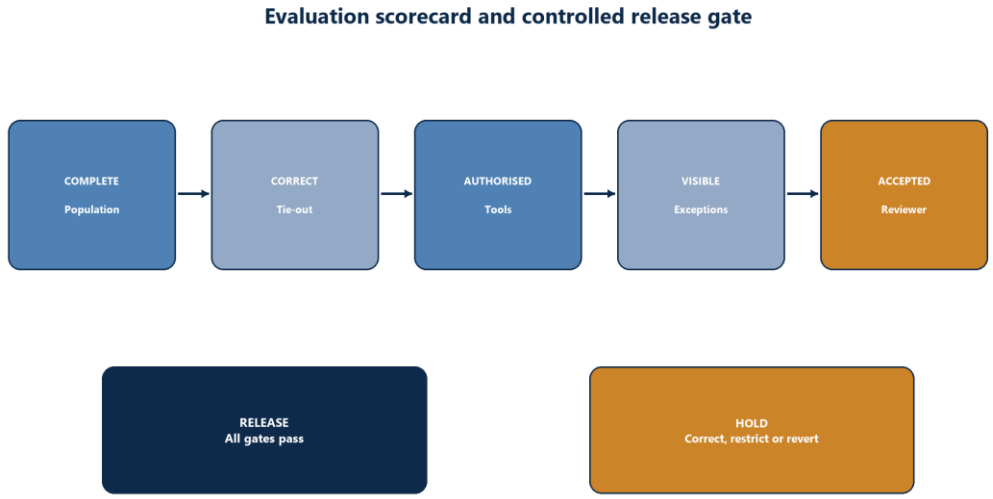

Figure 5. Evaluation scorecard and controlled release gate
Source: Matchpoint evaluation framework; risk and assurance context [7-18].

9. SECURITY, PRIVACY AND THIRD-PARTY RISK

9.1 Agent-specific threats

The NCSC secure-AI guidance recommends threat modelling across design, development, deployment, operation and maintenance [10]. OWASP identifies prompt injection, insecure output handling, sensitive-information disclosure and excessive agency among material LLM application risks [11]. MITRE ATLAS includes techniques for generative and agentic AI, including exfiltration and destructive tool invocation [12]. A back-office threat model should cover both conventional software risk and model-mediated paths.

Threat	Back-office example	Primary controls
Indirect prompt injection	Malicious instruction embedded in invoice or manager report	Treat content as data; isolate instructions; allowlist tools; approval
Excessive agency	Agent receives payment, posting or deletion rights	Least privilege; separate release adapter; dual control
Data exfiltration	Tool sends ledger or investor data to unauthorised endpoint	Egress allowlist; DLP; audience-bound tokens; monitoring
Insecure output handling	Generated formula or command executes downstream	Typed schemas; deterministic validation; no raw execution

Threat	Back-office example	Primary controls
Credential compromise	Service token reused across systems	Short-lived scoped identity; vault; rotation; audience validation
Tool substitution	Connector capability or description changes	Trusted registry; signed version; change alert; regression test
Trace leakage	Prompts/logs duplicate sensitive records	Data minimisation; access; retention and redaction
Denial/cost exhaustion	Loops or oversized documents consume resources	Turn, time, token and cost limits; circuit breaker

9.2 Personal data

The UAE Personal Data Protection Law applies to electronic processing within its scope and establishes controls, duties and data-subject rights [26]. The cited law includes a right to object to certain solely automated decisions. DIFC Regulation 10 addresses personal data processed through autonomous and semi-autonomous systems and defines deployer responsibility [27]. Saudi Arabia's PDPL defines processing broadly to include manual and automated operations [28]. UK ICO guidance distinguishes automated decisions from decision-support and emphasises meaningful human review [29]. The EU AI Act establishes harmonised requirements for systems within its material and territorial scope, including obligations that depend on the actor and risk classification [30].

An entity must obtain qualified advice on jurisdiction, role, lawful basis, transparency, rights, transfers and retention. The system design should record purpose and data category, minimise fields, restrict tools, document model/provider processing, honour retention, and avoid solely automated significant decisions unless specifically approved as lawful.

Privacy record	Required fields
Processing inventory	Purpose, data subjects, categories, systems, recipients and owner
Legal assessment	Applicable regime, role, basis, notice and rights process
Transfer record	Destination, provider, subprocessors and approved mechanism
Automated-decision assessment	Decision, significance, human role, contest and override
Retention	Source, packet, trace and deletion schedule
Incident	Data affected, access, chronology, containment and notification decision

9.3 Outsourcing and concentration

CBUAE's outsourcing regulation for banks requires risk management, oversight and continued ability to meet obligations, and addresses material business activities and supervisory access [23]. DORA requires in-scope EU financial entities to map ICT dependencies and manage third-party risk while retaining responsibility [31]. These regimes do not automatically apply to every B4 or A2 organisation. Their control questions remain relevant: what function depends on the provider, where is data processed, what are subcontractors, how is performance evidenced, what are audit rights, and how does the organisation exit?

Provider evidence	Minimum review
Service	Exact model, region, capacity, availability and support
Data	Inputs, outputs, logs, retention, training use and deletion
Security	Identity, encryption, testing, incidents and attestations
Resilience	Dependencies, recovery, status evidence and fallback
Contract	Confidentiality, IP, liability, audit, notice and termination
Portability	Export formats, prompts, tools, evals, embeddings and migration test
Concentration	Critical processes, spend, volume, alternatives and exit time

9.4 Cybersecurity governance

NIST CSF 2.0 organises cybersecurity outcomes around Govern, Identify, Protect, Detect, Respond and Recover [9]. A workflow owner should map the agent control plane, connected records, credentials, tools, models and providers into the organisation's cybersecurity programme. A separate dashboard records security incidents and near misses alongside operational quality.

10. REGULATORY DATA AND DIGITAL REPORTING CONTEXT

10.1 UAE eInvoicing

The UAE Ministry of Finance describes an eInvoice as structured invoice data exchanged electronically and reported to the Federal Tax Authority; PDF, Word, scanned image and email formats are not eInvoices [32]. The February 2026 guidelines announcement describes the national framework and operational expectations [33]. The portal and legislative material should be checked for current scope and dates before implementation.

Structured invoice status can become a source event in a reconciliation or compliance packet. It does not replace internal acceptance, delivery evidence, accounting treatment, payment status or dispute resolution. Each event remains distinct.

Invoice event	Source	Internal decision
Issued/received	Accredited exchange/provider record	Completeness and entity mapping
Validated/rejected	Network or authority status	Correct or route exception
Accepted/disputed	Contract and buyer process	Receivable/payable status
Accounted	ERP posting and accounting policy	Ledger recognition
Paid	Bank and payment records	Settlement and cash reconciliation
Reported	Tax return/workpaper	Tax-owner acceptance

The UAE FTA states that relevant corporate-tax records should be retained for at least seven years after the relevant tax period [34]. The exact record set and applicability require current tax advice. A packet can monitor presence, ownership and retention status; it cannot determine tax sufficiency by itself.

10.2 Saudi eInvoicing

ZATCA describes Phase 1 as generation and storage of compliant invoices and Phase 2 as integration with ZATCA systems in notified waves [35]. Its technical materials address structured formats, integration, data dictionaries and security requirements [36]. A GCC workflow should use jurisdiction-specific connectors and rules. UAE and Saudi status events cannot be treated as interchangeable.

10.3 Risk data and reporting

BCBS 239 sets principles for effective risk-data aggregation and risk reporting [19]. A January 2026 Basel newsletter states that accurate, comprehensive and timely aggregation and reporting remain critical and highlights continuing implementation challenges; it expressly says it does not create new supervisory guidance [20]. An A2 family office is generally outside the principles' original SIB scope, while a bank-owned or regulated structure may have direct obligations. The concepts of ownership, lineage, accuracy, completeness, timeliness and adaptability remain useful to consolidated reporting design.

10.4 Consumer and model data controls

CBUAE Consumer Protection Standards require in-scope licensed financial institutions to maintain data controls, authorised access and audit logs and to assign senior accountability [24]. CBUAE model-management data-governance requirements emphasise source identification, collection, quality, secure storage, infrastructure, ownership and independent validation for covered model data [25]. Application and scope require qualified regulatory mapping.

11. GOVERNANCE, ASSURANCE AND PROFESSIONAL BOUNDARIES

11.1 Three lines and accountable ownership

The first line owns the process, records, controls and accepted output. Risk and compliance functions set policy, monitor and challenge. Internal audit provides independent assurance under its charter. The 2024 Global Internal Audit Standards establish requirements for governance, management and performance of internal audit services [15]. The IIA AI Auditing Framework provides AI-specific governance and audit considerations [16].

Role	T18 accountability	Evidence
Board/owner committee	Risk appetite, material authority and resources	Approved charter and reporting
CFO/controller/CIO	Process outcome, definitions and release	Packet acceptance and exceptions
Technology owner	Architecture, access, reliability and changes	Service, test and incident records
Data owner	Source rights, quality and retention	Data inventory and quality results
Compliance/privacy	Policy, personal data and reserved decisions	Assessments and dispositions
Security	Threat model, monitoring and response	Tests, alerts and incident evidence
Internal audit	Independent risk-based assurance	Engagement work and conclusions
External adviser/auditor	Engagement-specific professional work	Independent report under applicable terms

11.2 COSO control design

COSO's Internal Control Integrated Framework is designed to improve confidence in data and information, and COSO lists 2026 guidance on effective internal control over generative AI [14]. A control catalogue should cover environment, risk assessment, control activities, information/communication and monitoring. Use of the framework does not establish that controls are effective; operating evidence and assurance are required.

11.3 Evidence available to auditors

A packet can make source lineage, transformations, access, exceptions and approvals easier to inspect. An auditor determines relevance, reliability, sufficiency and appropriateness under the applicable standards. PCAOB AS 1105 states that more evidence cannot compensate for poor-quality evidence and requires attention to information accuracy, completeness and precision [17]. Inquiry alone is not sufficient for relevant audit conclusions under the cited standard [17]. Agent-generated explanations therefore support, but do not replace, source evidence and audit procedures.

11.4 Authority matrix

Action	Initial agent authority	Human/professional authority
Read approved source	Yes, within purpose and scope	Owner approves access
Extract and normalise	Yes, with validation	Reviewer resolves ambiguity
Calculate controlled metric	Invoke deterministic service	Owner approves rule and changes
Draft narrative	Yes, with source links	Reporting owner accepts
Propose journal	Draft only	Qualified preparer/reviewer approves
Post journal	No	Delegated finance authority/system control
Propose payment	Draft packet only	Signatories approve and release
Clear sanctions/KYC alert	No	Compliance authority
File tax/regulatory report	No	Authorised officer/adviser
Make investment decision	No	CIO/investment committee
Delete source/record	No	Retention owner through controlled process

12. ILLUSTRATIVE PRODUCTIVITY AND ECONOMICS

12.1 Evidence boundary

No approved observed Matchpoint client result was supplied for this paper. Attributed revenue, cost reduction and loss reduction are USD 0. The following values are **unverified illustrative management assumptions**. They demonstrate arithmetic and measurement design only.

12.2 Illustrative monthly workflow

Assume a team produces 240 accepted reconciliation, reporting and compliance packets per month. The baseline averages 1.50 human hours per accepted packet, or 360 hours. The assisted workflow averages 0.65 preparation and review hours per accepted packet, or 156 hours, plus 58 monthly hours for exceptions, controls, support and governance, or 214 hours. The illustrative gross capacity released is 146 hours.

Input	Illustrative assumption	Status
Accepted packets per month	240	Unverified management assumption
Baseline hours per accepted packet	1.50	Unverified management assumption
Assisted preparation/review hours per accepted packet	0.65	Unverified management assumption
Monthly exception/control/support hours	58	Unverified management assumption
Baseline monthly hours	360	Calculated from assumptions
Assisted monthly hours	214	Calculated from assumptions
Gross hours released	146	Calculated; attributed value remains USD 0

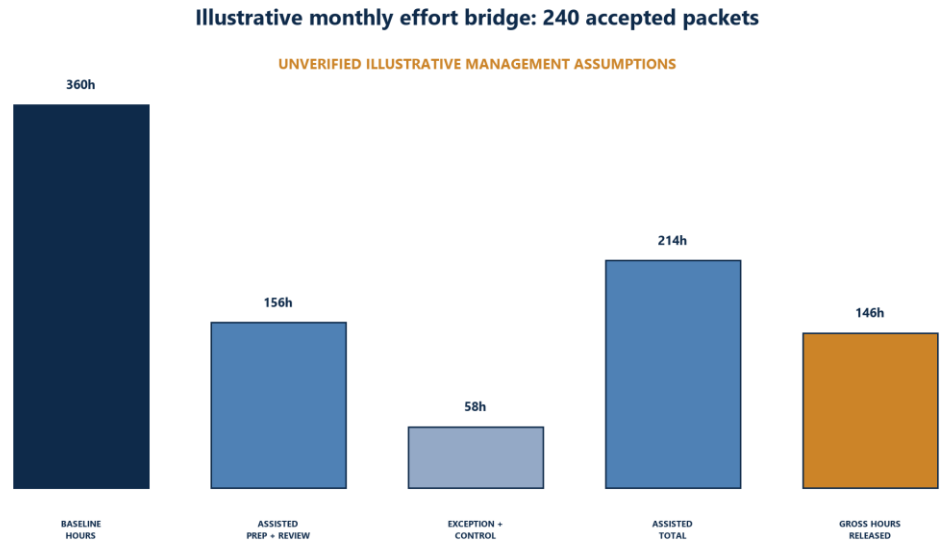

Figure 6. Illustrative productivity and full-cost bridge

Source: Unverified illustrative management assumptions; Matchpoint calculation.

The 146 hours are not an observed saving. They may become capacity, shorter cycle time or reduced paid cost. Any cost reduction requires a reconciled payroll or vendor-cost change and finance approval. Any revenue claim requires collected incremental revenue and an approved counterfactual. Any loss-reduction claim requires a risk-owner methodology.

12.3 Full-cost model

Annual full cost includes implementation, integrations, data remediation, security, licences, model and tool usage, monitoring, evaluation, reviewer time, vendor management, incidents, recovery and amortised change costs. Benefit lines remain separate.

Cost line	Measurement source	Common omission
Implementation	Project ledger and time records	Internal staff time
Data/integration	Engineering and provider invoices	Source remediation
Model/tool use	Provider billing and run ledger	Retries and failed packets
Human review	Time study by role	Exceptions and escalation
Control/assurance	Security, compliance and audit time	Periodic testing
Resilience	Fallback capacity and recovery tests	Standby/manual process
Change	Regression, approvals and migration	Provider/model updates
Exit	Export, rebuild and parallel run	Concentration and contract end

12.4 Benefit attribution

Benefit	Observation	Approval gate
Cycle-time reduction	Comparable trigger-to-acceptance records	Process owner confirms scope and quality
Capacity	Accepted packets per paid hour	Workload and service level comparable
Labour-cost reduction	Reconciled payroll/vendor change	Finance approval and causal link
Revenue	Collected incremental revenue	Customer, channel and counterfactual evidence
Working capital	Observed cash-date or balance change	Treasury/finance methodology
Loss reduction	Approved prevented-loss method	Risk and finance approval
Compliance quality	Severity-weighted errors and timeliness	Compliance owner and independent testing

Break-even occurs when approved annual benefits equal or exceed full annual cost. A claims register records period, source, counterfactual, owner and approval. Scenario values remain excluded from management accounts, investment materials and external claims.

12.5 Sensitivity

The dominant sensitivities are first-pass acceptance, reviewer time, exception incidence, change frequency, provider cost and failure severity. Higher generation speed can reduce value if rejection, rework or exception burden rises. The business case should therefore show volume, quality and control sensitivities together.

13. PROCUREMENT AND IMPLEMENTATION PLAYBOOK

13.1 Process discovery

The team maps the current process by observed events, records, decisions, roles, exceptions, elapsed time and effort. Interviews help identify variants. System logs, reconciliations, checklists and released reports provide stronger operating evidence. The current state is baselined before redesign.

Discovery artefact	Contents	Owner
Process map	Trigger, steps, records, decisions and outputs	Process owner
Record inventory	Systems, fields, owners, rights and quality	Data owners
Authority matrix	Prepare, review, approve, release and change	CFO/CIO/compliance
Exception taxonomy	Class, severity, threshold and disposition	Risk/process owner

Discovery artefact	Contents	Owner
Baseline study	Volume, cycle, touch, quality and cost	Finance/operations
Control register	Objective, activity, evidence, frequency and tester	Control owners

13.2 Build-or-buy diligence

The decision should compare process fit, control transparency, integration, data handling, evaluation access, provider concentration, skills, total cost and exit. A polished interface is insufficient evidence. Vendors should demonstrate a named workflow against representative and adverse cases in a controlled environment.

Diligence question	Required answer
Which model and tools act?	Named versions, routes, regions and change notice
What can the workflow write?	Exact systems, objects, fields and approval path
How are sources bound to claims?	Demonstrable lineage and immutable identifiers
How are calculations validated?	Deterministic services, tests and independent recomputation
What is retained or used for training?	Contractual and technical settings, subprocessors and deletion
How is failure handled?	Stop, alert, rollback, manual fallback and incident support
Can we evaluate independently?	Exportable cases, outputs, traces and metrics
How do we exit?	Data/config export, migration help, deletion and parallel run

13.3 Pilot charter

A pilot charter sets one workflow, one accountable sponsor, defined systems, fixed population, no prohibited action rights, gold and holdout sets, evaluation thresholds, incident process, reviewer capacity, budget, decision date and retirement condition. A pilot is a controlled test, not a permanent ungoverned production channel.

13.4 Training and operating change

Reviewers need task expertise, source-system knowledge, agent failure awareness, security escalation and authority discipline. They should know that fluent text is not evidence. Training includes adversarial cases, conflicting sources, unavailable tools, overconfident narratives and inappropriate action requests.

14. GATED 12-18 MONTH ROADMAP

The roadmap advances through evidence gates. Calendar periods are indicative; readiness determines movement. The initial stages can be shorter for a clean process or longer where records, ownership and controls need remediation.

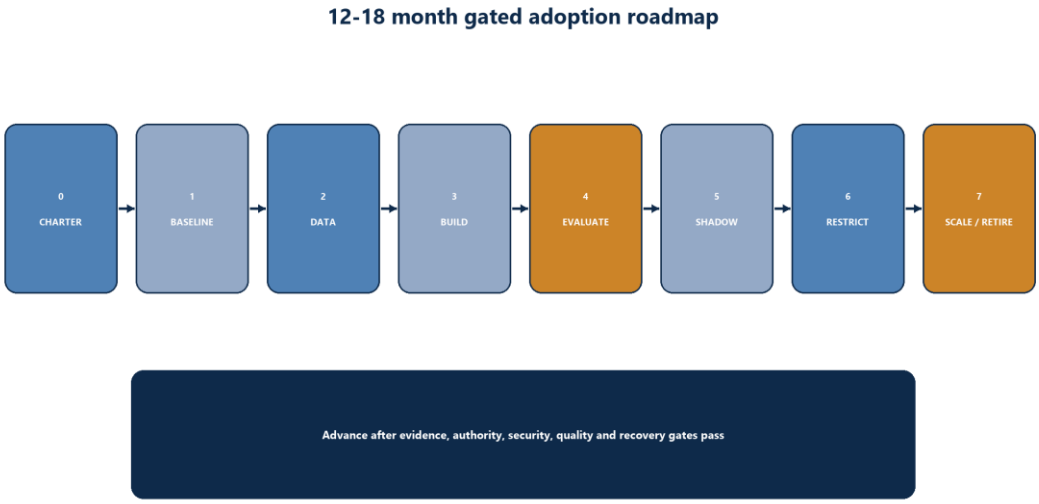

Figure 7. Gated 12-18 month adoption roadmap

Source: Matchpoint adoption framework.

Phase	Deliverable	Exit gate
0 Charter	Task contract, owner, value hypothesis and prohibited actions	Sponsor, authority and scope approved
1 Baseline	Current process, records, quality, time, effort and controls	Representative baseline accepted
2 Data	Source manifest, rights, schemas and quality remediation	Critical inputs complete and owned
3 Build	Deterministic services, bounded agent, ledger and review queue	Unit, integration and security tests pass
4 Evaluate	Gold, holdout, adverse and recovery results	Thresholds pass without hidden failures
5 Shadow	Live parallel packets and reviewer study	Stable acceptance, capacity and control evidence
6 Restricted	Named users, narrow limits and monitored release	Observed value and incidents within appetite
7 Scale/retire	Controlled expansion, redesign or archived evidence	Governance decision and maintained fallback

The first 90 days should end with a reproducible baseline and a testable packet, even if no production release occurs. Months four to nine establish representative evaluation, shadow use and provider controls. Months ten to eighteen can expand entities, accounts or packet types after observed stability. New action authority requires its own approval.

15. LIMITATIONS, RESEARCH AGENDA AND CONCLUSION

This paper is based on public sources available through 1 August 2026 and a conceptual operating framework. It does not report a named GCC client deployment, audited control result, regulatory approval, professional opinion or approved financial return. Laws, regulations, provider capabilities and technical specifications can change. Applicability requires current qualified review.

Empirical productivity evidence is task-specific. Customer support and software development studies do not establish back-office outcomes [2-4]. Public agent benchmarks do not replicate private systems, data, permissions, reviewers or consequences [1]. Vendor material describes intended designs and capabilities; it does not establish local operating effectiveness [5,6,13].

Further research should publish anonymised, representative evaluations for reconciliation, reporting and compliance packets; separate deterministic and model contributions; measure false positives, missed exceptions, reviewer effort and tail severity; report failed runs; and compare assisted, deterministic and manual baselines. GCC studies should

examine structured eInvoicing events, family-business entity complexity, bilingual documents, cross-border portfolio reporting and jurisdiction-specific data rights.

The operating recommendation is actionable. B4 owners and A2 family-office leaders should begin with one bounded workflow, define the accepted packet and authority ladder, repair source ownership, build deterministic controls, restrict tools, evaluate on representative cases, operate in shadow mode and measure full effort. The agent interprets, assembles and routes. Controlled systems calculate and record. Named professionals and delegated officers accept, release and remain accountable.

DECLARATIONS

Author. Prepared by Matchpoint Partners. Named-person author attribution is pending CK approval.
Evidence date. Public sources were reviewed through 1 August 2026.
Financial claims. No approved observed Matchpoint or client financial result was supplied. Attributed revenue, cost reduction and loss reduction remain USD 0. All worked figures, thresholds, times, rates and economics are unverified illustrative management assumptions.
Conflicts and funding. No external funding or paid sponsorship for this paper was identified in the supplied materials. This statement is based on the materials available to the authoring workflow and is subject to CK confirmation.
Professional boundary. This research paper is not investment, legal, regulatory, accounting, audit, tax, privacy, cybersecurity, valuation or technology advice. Readers should obtain qualified advice for their circumstances.

REFERENCES

[1] Stanford Institute for Human-Centered Artificial Intelligence (2026). AI Index Report 2026. https://hai.stanford.edu/ai-index/2026-ai-index-report
[2] Brynjolfsson, E., Li, D. and Raymond, L. R. (2023, revised 2023). Generative AI at Work. NBER Working Paper 31161. https://www.nber.org/papers/w31161
[3] Becker, J., Rush, N., Barnes, B. and Rein, D. (2025). Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity. METR. https://metr.org/blog/2025-07-10-early-2025-ai-experienced-os-dev-study/
[4] Becker, J., Rush, N., Cunningham, T., Rein, D. and Mahamud, K. (2026). We are Changing our Developer Productivity Experiment Design. METR. https://metr.org/blog/2026-02-24-uplift-update/
[5] OpenAI (2025). A Practical Guide to Building Agents. https://openai.com/business/guides-and-resources/a-practical-guide-to-building-ai-agents/
[6] Anthropic (2024). Building Effective Agents. https://www.anthropic.com/engineering/building-effective-agents
[7] National Institute of Standards and Technology (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). https://www.nist.gov/itl/ai-risk-management-framework
[8] National Institute of Standards and Technology (2024). Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile, NIST AI 600-1. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf
[9] National Institute of Standards and Technology (2024). The NIST Cybersecurity Framework (CSF) 2.0. https://www.nist.gov/publications/nist-cybersecurity-framework-csf-20
[10] UK National Cyber Security Centre and partners (2023). Guidelines for Secure AI System Development. https://www.ncsc.gov.uk/collection/guidelines-secure-ai-system-development
[11] OWASP Foundation (2025). OWASP Top 10 for LLM Applications v2.0. https://owasp.org/www-project-top-10-for-large-language-model-applications/assets/PDF/OWASP-Top-10-for-LLMs-v2025.pdf
[12] MITRE (current at 2026). MITRE ATLAS: Adversarial Threat Landscape for Artificial-Intelligence Systems. https://atlas.mitre.org/
[13] Model Context Protocol (2025). Specification and Security Best Practices. https://modelcontextprotocol.io/docs/tutorials/security/security_best_practices
[14] Committee of Sponsoring Organizations of the Treadway Commission (2026). Internal Control; Achieving Effective Internal Control Over Generative AI. https://www.coso.org/internal-control
[15] The Institute of Internal Auditors (2024). Global Internal Audit Standards. https://www.theiia.org/en/standards/2024-standards/global-internal-audit-standards/
[16] The Institute of Internal Auditors (2024). Artificial Intelligence Auditing Framework. https://www.theiia.org/en/content/tools/professional/2023/the-iias-updated-ai-auditing-framework/
[17] Public Company Accounting Oversight Board (current at 2026). AS 1105: Audit Evidence. https://pcaobus.org/oversight/standards/auditing-standards/details/AS1105
[18] International Auditing and Assurance Standards Board (2026). Audit Evidence and Risk Response; ISA 330, ISA 500 and ISA 520 Project. Exposure-draft status at 1 August 2026. https://www.iaasb.org/consultations-projects/audit-evidence-and-risk-response-isa-330-isa-500-isa-520
[19] Basel Committee on Banking Supervision (2013). Principles for Effective Risk Data Aggregation and Risk Reporting. https://www.bis.org/publ/bcbs239.htm
[20] Basel Committee on Banking Supervision (2026). Implementation of the BCBS 239 Principles. Informational newsletter; no new supervisory guidance. https://www.bis.org/publ/bcbs_nl36.htm
[21] Basel Committee on Banking Supervision (2021). Principles for Operational Resilience. https://www.bis.org/bcbs/publ/d516.htm
[22] Central Bank of the United Arab Emirates (in force; accessed 2026). Operational Risk Standards. https://rulebook.centralbank.ae/en/rulebook/operational-risk-standards
[23] Central Bank of the United Arab Emirates (in force; accessed 2026). Outsourcing Regulation for Banks. https://rulebook.centralbank.ae/en/rulebook/outsourcing-regulation-banks
[24] Central Bank of the United Arab Emirates (in force; accessed 2026). Consumer Protection Standards. https://rulebook.centralbank.ae/en/rulebook/consumer-protection-standards
[25] Central Bank of the United Arab Emirates (in force; accessed 2026). Model Management Standards, 5.1 Data Governance. https://rulebook.centralbank.ae/en/rulebook/51-data-governance
[26] United Arab Emirates Government (2021; accessed 2026). Federal Decree-Law No. 45 of 2021 Regarding the Protection of Personal Data. https://u.ae/en/about-the-uae/digital-uae/data/data-protection-laws
[27] Dubai International Financial Centre (current at 2026). Data Protection Regulations, Regulation 10: Personal Data Processed through Autonomous and Semi-Autonomous Systems. https://assets.difc.com/v1/media/edge/images/dubaiintern0078-difcexperie96c5-production-3253/media/project/difcexperiences/difc/difcwebsite/documents/laws--regulations/data-protection-regulation.pdf
[28] Saudi Data and Artificial Intelligence Authority (current at 2026). Personal Data Protection Law. https://sdaia.gov.sa/en/SDAIA/about/Documents/Personal%20Data%20English%20V2-23April2023-%20Reviewed-.pdf
[29] UK Information Commissioner's Office (current at 2026). Guidance on AI and Data Protection; Individual Rights and Automated Decision-Making. https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/how-do-we-ensure-individual-rights-in-our-ai-systems/
[30] European Union (2024). Regulation (EU) 2024/1689 Laying Down Harmonised Rules on Artificial Intelligence. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32024R1689
[31] European Union (2022). Regulation (EU) 2022/2554 on Digital Operational Resilience for the Financial Sector. https://eur-lex.europa.eu/eli/reg/2022/2554/oj
[32] UAE Ministry of Finance (current at 2026). UAE eInvoicing Programme. https://mof.gov.ae/en/about-us/initiatives/einvoicing/

[33] UAE Ministry of Finance (2026). Ministry of Finance Issues UAE Electronic Invoicing Guidelines to Support National Rollout. https://mof.gov.ae/en/news/ministry-of-finance-issues-uae-electronic-invoicing-guidelines-to-support-national-rollout/
[34] UAE Federal Tax Authority (2025). Record and Documentation Retention for Corporate Tax. https://tax.gov.ae/en/media.centre/news/pr.28082025.aspx
[35] Zakat, Tax and Customs Authority, Saudi Arabia (current at 2026). E-Invoicing Roll-Out Phases. https://zatca.gov.sa/en/E-Invoicing/Introduction/Pages/Roll-out-phases.aspx
[36] Zakat, Tax and Customs Authority, Saudi Arabia (current at 2026). E-Invoicing Detailed Technical Guidelines and Supporting Documents. https://zatca.gov.sa/en/E-Invoicing/Introduction/Guidelines/Pages/default.aspx
[37] Financial Action Task Force (2021). Opportunities and Challenges of New Technologies for AML/CFT. https://www.fatf-gafi.org/en/publications/Digitaltransformation/Opportunities-challenges-new-technologies-for-aml-cft.html
[38] US Department of the Treasury, Office of Foreign Assets Control (2019). A Framework for OFAC Compliance Commitments. https://ofac.treasury.gov/media/16331/download
[39] Federal Financial Institutions Examination Council (current at 2026). BSA/AML Examination Manual; Suspicious Activity Reporting and Automated Account Monitoring. https://bsaaml.ffiec.gov/manual/AssessingComplianceWithBSARegulatoryRequirements/04
[40] IFRS Foundation (2026). Using the IFRS Digital Taxonomies; The Taxonomy Architecture. https://www.ifrs.org/content/dam/ifrs/standards/taxonomy/general-resources/ifrs-taxonomy-architecture/taxonomy-architecture-2026.pdf

APPENDIX A. MINIMUM TASK CONTRACT

Field	Required content
Identity	Workflow, entity, period, population and owner
Objective	Decision supported and accepted output
Sources	Systems, objects, cut-off, completeness and rights
Rules	Deterministic calculations, classifications and thresholds
Agent	Model, tools, instructions, memory and limits
Exceptions	Classes, severity, stop and escalation
Authority	Prepare, review, accept, release and change roles
Quality	Gold/holdout sets, metrics, thresholds and reviewer
Security	Threat model, identity, scopes, logs and incident process
Recovery	Timeout, fallback, last safe state and owner
Retention	Source, packet, trace, approval and deletion schedule

APPENDIX B. PACKET SCHEMA

Block	Core fields
Header	Packet ID, task contract, state, entity, period and timestamps
Manifest	Source IDs, hashes, schema, owners, quality and rights
Transform	Code/rule/model/tool versions and parameters
Results	Deterministic values, generated narrative and limitations
Exceptions	Identity, severity, evidence, owner and disposition
Validation	Test IDs, outcomes, independent checks and release gate
Acceptance	Reviewer, role, timestamp, decision and limitations
Release	Target, adapter, idempotency, response and reconciliation

APPENDIX C. MINIMUM CONTROL REGISTER

Control objective	Example activity	Evidence
Complete population	Count, total and sequence check at ingest	Manifest and validator result
Correct calculation	Independent deterministic recomputation	Rule version and comparison
Authorised access	Purpose-bound role and service identity	Access decision and logs
Restricted action	Read/write separation and approval adapter	Tool register and attempted-action tests
Supported narrative	Claim-to-source binding	Claim register and source IDs
Visible exceptions	Fail-closed queue with ownership	Exception ledger and ageing
Approved release	Authenticated acceptance and dual control	Acceptance and release event
Recoverable service	Tested fallback and last safe state	Recovery test and result
Controlled change	Regression and approval before release	Change record and test pack
Retained evidence	Policy-linked archive and deletion	Retention log

APPENDIX D. ILLUSTRATIVE FORMULAE

Baseline hours = accepted packets x baseline hours per accepted packet.
Assisted hours = accepted packets x assisted preparation/review hours per accepted packet + exception/control/support hours.
Gross hours released = baseline hours - assisted hours.
First-pass acceptance = packets accepted without correction / reviewed packets.
End-to-end cycle time = accepted timestamp - task trigger timestamp.
Full annual cost = implementation amortisation + data/integration + model/tool + review + control/assurance + resilience + change + exit provision.

Approved annual benefit = approved labour-cost reduction + approved operating-cost reduction + approved working-capital benefit + approved loss reduction + approved collected incremental revenue.
All numerical inputs require observed evidence and named approval before commercial attribution.

APPENDIX E. RED-FLAG REGISTER

Red flag	Why it matters	Immediate response
Agent can post, pay, file or delete with its preparation identity	Excessive authority and weak segregation	Disable action; review access and incidents
Output lacks immutable source identifiers	Claims cannot be reproduced	Reject packet and repair lineage
Generated narrative changes a controlled calculation	Model has crossed the deterministic boundary	Reject; isolate calculation service
Reviewer sees only summary text	Evidence and limitations are hidden	Redesign review view
Success rate excludes abandoned runs	Performance is overstated	Reconstruct full denominator
Provider/model changed without regression	Quality and security state is unknown	Suspend release and retest
Prompt or tool description arrives inside business content	Injection path exists	Treat as data; quarantine and test
Logs contain uncontrolled personal or confidential data	Secondary data store creates exposure	Restrict, minimise and remediate retention
“Hours saved” lacks comparable baseline and review effort	Value is unverified	Set attributed benefit to zero
Compliance or tax decision has no named professional owner	Reserved judgement is displaced	Stop and escalate