Acquiring AI Agents Liability Permissions and Human Override Diligence

A Transaction Framework for Delegated Authority Consequential Actions Control Evidence and Deal Protection

Authored by Chennakeshav Adya

Independent Researcher

September 2026

Abstract

An acquisition of an agentic artificial-intelligence business transfers more than software. It can transfer a system that receives authority, uses credentials, selects tools and changes the legal, financial or operational state of customers and the combined group. A conventional technology review may identify models, repositories and security controls while leaving the buyer unable to answer five transaction questions: who authorised each action, which permissions made it possible, which person retained decision authority, what happened when the system failed, and who bears the resulting cost.

This paper develops an acquisition-diligence framework for liability, permissions and human override. It maps an agent from principal and purpose through identity, delegated authority, tool call, consequential action, approval, execution, evidence and remedy. It separates read, recommend, prepare, approve and execute permissions; classifies actions by financial consequence, legal effect, external visibility and reversibility; and tests whether a human control is informed, timely, independent and technically effective. The framework connects operational evidence to valuation, representations, indemnities, escrow, insurance, closing conditions, integration sequencing and contingent consideration.

Four hypothetical acquisition cases illustrate the method: a treasury-payment agent, a customer-contract agent, a cloud-administration agent and a regulated-claims agent. Management assumptions produce annual expected loss and control-remediation estimates that vary materially with privilege, action volume, detection time and recovery. A separate probability-weighted liability illustration produces an expected exposure of USD 29.4 million before insurance, indemnity recovery and tax. These values explain the framework. They are not observations about a company, forecast or valuation conclusion.

The analysis finds that a generic human-in-the-loop statement is weak transaction evidence. A buyer needs the actual production permission graph, policy enforcement point, approval record, signed action trace, incident history, customer allocation of responsibility and tested stop or recovery mechanism. Value should be released as the buyer verifies bounded authority, reproducible controls, transferable rights, measurable residual exposure and a post-close operating model that preserves accountable human decision making.

JEL Classification: G24, G34, K12, K13, L86, M15, O33

Keywords: agentic AI M&A, AI liability, agent permissions, human override, delegated authority, AI due diligence, technology transactions, consequential actions

Introduction

Agentic AI changes the diligence perimeter because software can move from producing information to initiating action. An agent may draft a payment, create a user, alter cloud infrastructure, communicate with a customer, commit inventory, approve a refund, submit a filing or prepare a regulated conclusion. The

same model can be low risk in a read-only research workflow and transaction-critical when connected to credentials and production tools. The acquisition team therefore needs to diligence authority and consequence at the workflow level.

The current standards landscape is developing. NIST launched an AI Agent Standards Initiative in February 2026 and separately proposed work on software and AI-agent identity and authorisation [7-10]. Its published materials identify identification, authorisation, auditing, non-repudiation and prompt-injection controls as practical issues. OWASP describes excessive agency as a combination of excessive functionality, permission or autonomy and recommends least privilege, downstream authorisation and human approval for high-impact actions [17-19]. These sources define useful control questions. They do not establish that a target has implemented the controls or that any particular allocation of liability is legally effective.

Human oversight also has to operate in fact. Article 14 of the European Union AI Act requires certain high-risk systems to be designed for effective oversight, including the ability to disregard, override, reverse or interrupt outputs where appropriate [23-24]. A displayed approval button does not show that a reviewer understood the action, had sufficient information, held appropriate authority or could prevent execution. Transaction diligence must test the full path from request to consequence and remedy.

This paper provides that path. It treats liability as an economic exposure requiring legal advice, and permissions as a technical and organisational system requiring security evidence. It then connects both to price, deal protection and integration. The framework is designed for strategic buyers, private-equity sponsors, boards, transaction teams and lenders assessing an AI-agent business or an AI-enabled target.

1 Define the acquisition decision and liability perimeter

The first document should state why the buyer is acquiring the business and which agentic capability is included in value. The thesis may depend on customer distribution, proprietary workflow data, lower service cost, a specialist team, automation of the buyer's operations or ownership of a control layer. Each thesis creates different liability and integration questions. A product acquired for internal deployment can expose the buyer directly to employee, customer and regulatory consequences that were limited while the target operated as a vendor.

The perimeter should identify every legal entity, product, agent, workflow, environment, customer configuration, model, connector, tool, credential, service account, policy engine, approval service, data store and incident process. It should distinguish components owned by the target from customer-controlled identities, open-source software, third-party models and partner platforms. A demonstration can show an action without showing whose authority was used or whether that authority transfers at closing.

The buyer should list consequential actions before reviewing generic architecture. Consequence includes movement of money, creation or termination of rights, disclosure of information, changes to production systems, statements to customers or regulators, employment decisions, safety effects and irreversible publication. The list becomes the organising record for legal, security, financial, operational and insurance diligence.

Liability should be separated into observed obligations, asserted claims, possible obligations and prospective operating risk. IAS 37 defines principles for provisions and contingent liabilities, while IFRS 3 addresses liabilities assumed in a business combination [1-3]. The accounting conclusion is transaction-specific. The diligence record should preserve facts, ranges, uncertainty and responsible advisers rather than substitute a technical score for legal or accounting analysis.

2 Map the principal agent authority chain

Every production action should be traceable through a principal-agent-authority chain. The principal is the person or organisation whose objective is being pursued. The agent is the software instance acting within a defined workflow. Authority is the bounded set of actions the principal or another authorised party has delegated. The chain also includes the identity used, policy applied, tool selected, parameters supplied, approval received, execution response and retained evidence.

The target should demonstrate how a user's authority is represented. An agent acting for an employee through that employee's OAuth scope differs from an agent using a shared privileged service account. The shared account may be operationally convenient while weakening attribution, segregation and revocation. The buyer should test whether delegated authority remains bounded by user, purpose, resource, time, value and action type.

Authority can be lost in orchestration. A coordinating agent may delegate to specialist agents that call tools through another platform. The final system may record only the last tool call. Diligence should reconstruct the full delegation chain and verify that a downstream component cannot obtain greater authority than the originating principal. The control should be enforced by a trusted system, not solely by instructions interpreted by a model.

Non-repudiation matters when the action creates a dispute. The record should show which authenticated identity requested the work, which agent and version operated, which evidence was presented, which human approved, what exact action was executed and whether the record can be altered. NIST's current agent-identity work highlights auditing and non-repudiation as areas for implementation attention [8-10]. The buyer should inspect the target's implementation and test representative records.

3 Build the production permission graph

A permission inventory is insufficient when it lists accounts without connecting them to actions. The production permission graph should connect principals, agents, tools, resources, functions, environments and approval rules. It should show both explicit permissions and effective access inherited through groups, roles, tokens, platforms and customer integrations. Temporary privileges, emergency roles and dormant connectors belong in the same graph.

Permissions should be classified by operation: discover, read, create, modify, delete, approve, execute, export, impersonate, delegate and administer. The graph should distinguish production from test and internal from customer environments. It should also identify whether a credential can be used outside the intended agent path. A narrowly described tool can still run under an account with broad database, cloud or messaging permissions.

The buyer should compare required and effective permissions. Required permission follows the defined task and customer promise. Effective permission is what the identity and underlying system actually allow. The difference is permission surplus. Surplus should be quantified by reachable resources, action types and potential consequence. A target may reduce interface options while retaining broad downstream authority, leaving the blast radius unchanged.

Evidence should include current identity-provider exports, cloud and application roles, tool definitions, connector configurations, policy-as-code, token lifetimes, secret stores, approval rules and revocation tests. Screenshots and policy documents provide context. Machine-readable exports and observed execution provide stronger evidence of the production state.

4 Classify actions by consequence and reversibility

Action classification determines which controls and deal responses are proportionate. A useful matrix scores financial consequence, legal effect, information sensitivity, operational disruption, external

visibility, affected population and reversibility. The score should be based on the action's plausible result, including chained actions, rather than the apparent simplicity of the tool call.

Reversibility requires more than an opposite command. A deleted record may be restored while an external disclosure cannot be recalled. A payment can sometimes be reversed after funds move, subject to delay and recovery risk. A customer message can be corrected while contractual reliance or reputational harm remains. The classification should therefore distinguish technically reversible, operationally recoverable, financially recoverable and effectively irreversible actions.

Action velocity affects exposure. A weak decision executed once under review differs from the same weakness executed across thousands of accounts before detection. Limits on value, volume, frequency and customer population can bound loss. The buyer should test whether limits apply at the policy enforcement point and whether an agent can split activity across multiple calls to evade them.

The classification should drive approval, logging, monitoring, recovery and insurance analysis. Read-only evidence gathering may operate under automated controls. Preparation of a consequential action may require validation. Execution of an irreversible or externally visible action may require an independent authorised person. Actual legal requirements depend on the workflow and jurisdiction.

5 Test policy enforcement outside the model

Natural-language instructions can guide an agent. They should not be the sole mechanism authorising a consequential action. The buyer should identify the policy enforcement point that validates each tool request against identity, resource, action, context and limits. Complete mediation means every relevant request is checked, including retries and calls made through secondary agents or cached credentials.

OWASP's excessive-agency guidance recommends limiting functionality, permissions and autonomy and enforcing authorisation in downstream systems [17]. The acquisition team should test these controls through adversarial and ordinary scenarios. It should attempt disallowed resources, modified parameters, indirect prompt injection, expired authority, conflicting instructions, repeated calls and requests that cross value or volume limits.

Policy change is also consequential. The target should show who can modify tool schemas, allowed actions, approval thresholds, system prompts, model routing and service-account roles. Changes should follow segregation, review, testing, release and rollback. A developer who can modify both agent logic and production policy may be able to bypass a nominal approval control.

The buyer should retain test evidence linking the policy version to the production version. A laboratory control is useful only if the same enforcement path operates for customer work. Exceptions, break-glass access and manual overrides should have defined authority, limited duration, enhanced logging and retrospective review.

6 Distinguish recommendation preparation approval and execution

Transaction teams should avoid a binary classification of human or autonomous. A workflow can observe evidence, recommend an action, prepare a transaction, request approval, execute an approved transaction and verify the result. Different components may own each stage. The diligence map should identify where information becomes a decision and where a decision becomes a change in the world.

A recommendation can still create liability when it is presented as definitive, relied upon systematically or generated from unauthorised data. Preparation can create risk when it populates a payment, contract or configuration that reviewers approve routinely. Approval can be weak when the reviewer sees only a summary. Execution can depart from the approved action if parameters or state change after approval.

The buyer should test binding between approval and execution. The approval record should identify the exact action, resource, amount, counterparty, policy and expiry. Material changes should invalidate the approval. Execution should reject stale, altered or replayed approvals. The system should retain both proposed and executed states and reconcile differences.

This separation also informs valuation. A product that reliably prepares work under human authority may deliver substantial value without autonomous execution. Management forecasts should not assume that removing approval increases value when customers, regulators or insurers require accountability. The economic model should include the cost and throughput of the control design actually accepted by customers.

7 Evaluate whether human override is effective

A human-in-the-loop label should be decomposed into capability, information, authority, timing, independence and workload. The reviewer needs enough information to understand the action and its consequences. The reviewer must have actual authority to refuse, modify or stop it. The intervention must occur before irreversible execution or early enough to contain harm.

Interface design matters. The approval display should distinguish model-generated explanation from source evidence, show key parameters and highlight policy exceptions. It should avoid manipulative defaults and ambiguous bundles. Approval fatigue can turn a nominal control into routine confirmation. The target should measure review volume, time, rejection, modification, escalation and downstream outcomes.

Independence depends on the action. A user may approve an ordinary workflow while a financial, security or regulated action requires a different role. Segregation of duties should be implemented in identity and policy, not merely documented. The agent should not be able to select its own reviewer, suppress inconvenient evidence or rewrite the approval request after a human decision.

Override and stop mechanisms should be tested under load and failure. The buyer should observe revocation of an agent's credential, interruption of queued actions, containment of in-flight work and safe recovery. Article 14 of the EU AI Act identifies monitoring, interpretation, override, reversal and interruption as relevant oversight capabilities for certain high-risk systems [23]. Applicability requires legal analysis; the operational tests remain useful across transactions.

8 Reconstruct incidents near misses and hidden intervention

Incident diligence should include unauthorised action, excessive permission, prompt injection, data exposure, incorrect decision, failed approval, replay, tool misuse, runaway cost, customer complaint and control bypass. Near misses and manual recoveries are valuable because they show where the system almost created a consequence or depended on undocumented intervention.

The buyer should reconcile several records: security tickets, customer support, service credits, engineering issues, model-evaluation failures, cloud logs, insurance notices, legal claims, refunds and board reporting. A single register may omit events classified as product quality, customer success or operational error. Common identifiers and timeline analysis can reveal related events.

For each event, diligence should identify initiating condition, permissions used, detection time, affected scope, containment, recovery, customer communication, cost, legal assessment and remediation validation. Root-cause labels should distinguish model behaviour, data, orchestration, tool, identity, policy, interface, human review and organisational process.

Hidden intervention affects both risk and economics. Specialists may monitor agents continuously, repair actions and calm customers without appearing in product metrics. The buyer should sample workflows from request to final outcome and reconcile human time to payroll and support systems. An acquisition

model that removes this labour before controls are redesigned can increase liability while overstating synergy.

9 Quantify exposure through action cohorts

Expected exposure should be estimated by action cohort rather than a single company-wide probability. Cohorts can be defined by workflow, action class, customer, jurisdiction, permission level, approval design, model version and environment. Each cohort should have observed volume, exception, unauthorised attempt, override, reversal, incident and recovery measures.

A simplified expected-loss model multiplies action volume by event probability and consequence, then adds detection, response, customer, legal, regulatory and remediation cost. Tail scenarios require separate treatment because historical frequency may be low and impact concentrated. Correlation matters when a shared policy, credential, model or connector affects many customers simultaneously.

Management assumptions should be explicit and kept separate from observed evidence. A lack of claims does not establish a low probability when deployment history is short, permissions expanded recently or incidents were not captured. External benchmarks may inform scenario design but seldom match the target's workflow, controls and contractual allocation.

The model should show gross exposure, insurance assumptions, contractual caps, indemnity rights, recoverability and residual exposure. Recoverability should be tested for exclusions, limits, notice, retention, counterparty credit and timing. The acquisition decision should remain robust if recovery is delayed or disputed.

10 Review customer contracts and responsibility allocation

Customer contracts should be mapped to the actual workflow. Relevant clauses include service description, permitted use, customer instructions, approval, data roles, model and subprocessor changes, security, audit, incident notification, warranties, disclaimers, service levels, indemnities, liability caps, insurance, termination and transition. The buyer should compare negotiated exceptions across customers.

Responsibility may be divided among provider, customer, model vendor, cloud platform, integration partner and end user. A clause stating that the customer remains responsible for decisions may have limited practical value if the product executes actions without meaningful customer control or the sales process represented a managed outcome. Legal counsel should assess enforceability and conduct in each jurisdiction.

The diligence team should reconcile contractual permissions to technical permissions. If the contract authorises read-only access while production credentials allow modification, the gap is transaction-critical. If customers require approval before model changes, the target should show how versions and notices are managed. Undocumented customer-specific controls can delay integration.

Revenue quality is linked to liability allocation. Outcome pricing can increase willingness to pay while transferring operating risk to the provider. Minimum commitments can support cash while customers reserve broad termination or service-credit rights. The model should include the delivery and control cost required by the actual promise.

11 Assess regulatory and jurisdictional pathways

Agentic AI may intersect with sector, consumer, employment, financial, privacy, cybersecurity, product, competition and professional rules. The transaction team should map each consequential workflow to the legal entity providing it, customer location, affected person, data location and decision type. A global product can have different obligations and risk allocations across deployments.

The EU AI Act establishes a risk-based framework and includes human-oversight requirements for high-risk systems [23-24]. Data-protection authorities publish guidance on automated decision making, accountability and data governance [25-26]. US federal and state requirements continue to develop, while sector regulators can apply existing laws to AI-enabled conduct. The paper does not provide a legal conclusion on applicability.

The buyer should request the target's legal inventory, classification analyses, impact assessments, regulator correspondence, customer representations and change-monitoring process. It should test whether the inventory is tied to the current product and jurisdictions. A generic policy can be outdated when workflows, permissions or customer populations change.

Regulatory change belongs in valuation and integration. Compliance work may require new approval roles, data controls, documentation, testing, customer notices or product limitations. The transaction model should include cost, timing and revenue effects. Closing conditions may be appropriate where lawful operation or transfer depends on a material consent or remediation.

12 Test data privacy intellectual property and confidentiality

Permissions to act often imply permissions to access data. The buyer should map data sources, purpose, legal basis, customer instruction, retention, transfer, model use and deletion for each workflow. Agent memory, traces and observability can replicate sensitive information beyond the primary system. Tool results can introduce data from a source the user was not authorised to access.

The rights register should cover code, prompts, policies, tool schemas, workflow designs, evaluation sets, customer configurations, training and feedback data, documentation, patents, trademarks and trade secrets. It should identify creator, assignment, licence, restriction, sublicensing, change of control and termination. Customer data access needed to perform a service does not automatically create a transferable asset.

Confidentiality risk extends to actions. An agent may send information, populate an external system or disclose reasoning through an approval interface. The buyer should test destination allowlists, data-loss prevention, redaction, customer segregation and logging. It should verify that controls apply to retries, secondary agents and support tooling.

IFRS 3 and IAS 38 provide accounting requirements relevant to identifiable intangible assets in a business combination [1,4]. Purchase-price allocation does not resolve whether the buyer has operational authority to use customer data or third-party technology. The legal and technical transfer analysis should precede valuation assumptions.

13 Evaluate third party models tools and agent protocols

Agentic products commonly depend on model providers, clouds, identity systems, enterprise applications, data services and connectors. The buyer should inventory each dependency, agreement, permission scope, service level, price, data treatment, audit right, change control, continuity, indemnity and termination. A material dependency may sit behind a tool or protocol rather than appear in the target's primary architecture.

Protocol interoperability can increase distribution and also widen the permission surface. The buyer should identify how servers and tools are discovered, authenticated, described and trusted. Tool metadata can influence model selection. Updates to a connector or schema can change effective behaviour without a model change. Registries, signing, allowlists, version pinning and testing should be inspected.

Provider terms can allocate responsibility to the target even when failure originates in a third-party service. The target may then owe a customer more than it can recover. The transaction model should show the chain of caps, exclusions and insurance. Concentration should include common model, cloud, identity and connector dependencies across customers.

Portability claims require a production-like test. Substituting a model or tool can change behaviour, latency, cost, evaluation and customer approval. The buyer should measure the transition and any period of restricted functionality. A multi-provider diagram without tested substitution is weak evidence of resilience.

14 Examine security prompt injection and confused deputy risk

An agent can become a confused deputy when it uses legitimate authority to pursue an unauthorised objective. Indirect prompt injection, manipulated tool output, compromised memory or a malicious peer agent can influence actions. The consequence depends on effective permissions and policy enforcement. Security review should therefore connect attack path to business action and recovery.

The buyer should inspect threat models, red-team results, security tests, dependency scanning, secrets management, isolation, monitoring and incident response. It should test inputs from customer documents, messages, websites and tool outputs. It should also test whether approval content can be manipulated so that a reviewer authorises a different action from the one executed.

Least privilege reduces potential harm. Separate identities for agent, user and service can improve attribution when designed correctly. Short-lived credentials, resource restrictions, transaction limits, destination controls and independent verification can bound exposure. Logging should preserve enough context to reconstruct the event without creating an uncontrolled store of confidential data.

MITRE ATLAS and OWASP provide threat and control taxonomies that can structure testing [17-21]. The target's risk should be assessed through its own workflow, architecture and customer commitments. Passing a generic checklist cannot demonstrate that production actions are authorised and recoverable.

15 Connect liability to valuation and purchase price

Traditional valuation methods remain relevant, including discounted cash flow, market approaches, precedent transactions and cost approaches [5-6]. Agentic liability affects inputs through revenue durability, contribution, control cost, insurance, working capital, remediation, customer retention, integration timing and tail exposure. The buyer should avoid applying a high-growth multiple before reconstructing these items.

The standalone forecast should include the operating control model customers and regulators will accept. Human review, security operations, evaluation, legal support and incident readiness are delivery costs. Removing them as immediate synergies can overstate value. Additional controls may improve conversion and retention, but that benefit should be supported by evidence and separately modelled.

The liability schedule should distinguish known obligations, specific contingent matters, control remediation and future operating risk. Known items may affect net debt or working capital depending on the agreement and accounting. Uncertain matters may influence price, escrow, indemnity or insurance. Prospective risk may affect the business plan and integration rather than create an acquisition-date liability.

Value should be released by evidence state. A target with bounded permissions, tested human controls, complete action records, stable incident performance and aligned contracts supports a stronger forecast than one relying on policy statements. The probability-weighted model should show how each unresolved control changes cash, timing and downside.

16 Design representations indemnities escrow and insurance

Transaction documents can allocate identified risk when definitions reflect the technical reality. Representations may address authority, customer instructions, data use, intellectual property, security,

incidents, regulatory compliance, material model or tool changes, evaluation records and insurance. Disclosure schedules should identify exceptions at workflow and customer level.

Specific indemnities can address known claims or defined exposures. Escrow or holdback can support recoverability. Warranty and indemnity insurance may transfer selected representation risk subject to exclusions, retention and underwriting. Cyber, technology errors and omissions, professional indemnity and other policies should be reviewed for insured entity, period, trigger, exclusions, limits, sublimits and notice.

Insurance should not be modelled at face value. Coverage for algorithmic error, autonomous action, regulatory penalty, contractual liability or known circumstance may be limited or excluded. The buyer should review policy wording, broker submissions, claims history and change-of-control treatment with advisers. Residual exposure remains after limits, retention and collection delay.

The deal team should align legal protection with action cohorts. A broad warranty may be difficult to prove and recover. A defined representation tied to production permission exports, incident records and customer exceptions can be more testable. Drafting and enforceability require transaction counsel.

17 Use closing conditions and covenants to remediate control gaps

Closing conditions should focus on matters necessary to transfer and operate the acquired business. Examples include revoking orphaned credentials, reducing critical permission surplus, implementing approval binding, preserving logs, assigning intellectual property, obtaining material customer or vendor consents and resolving a severe incident. Evidence and acceptance tests should be defined before signing where possible.

Lower-severity work can be managed through pre-close covenants, integration plans and specific budgets. The seller should maintain ordinary-course controls and notify material changes to models, tools, permissions, incidents and customer obligations. The buyer should avoid assuming operational control before closing in a way that creates legal or competition concerns.

Deferred consideration can be linked to durable operating evidence. Measures may include closure of critical findings, verified customer retention, adjusted contribution after control cost, tested portability and completion of agreed integration milestones. Incentives should not reward increased autonomous volume without acceptance, safety and cash.

The transaction file should show which gap affects decision, price, structure, timing or integration. A long list without materiality can delay the deal while failing to resolve the highest-consequence paths. Governance should assign an owner, evidence standard and deadline to each condition.

18 Construct four hypothetical acquisition cases

The treasury-payment agent prepares and, within limits, executes supplier payments. Its key exposures are payment authority, beneficiary change, value limits, segregation, replay and recovery. The customer-contract agent drafts and sends amendments within approved templates. Its exposures include apparent authority, unauthorised commitments, disclosure, version control and customer reliance.

The cloud-administration agent diagnoses incidents and changes infrastructure. Its exposures include privilege escalation, security configuration, service disruption, data access and cascading changes. The regulated-claims agent assembles evidence and recommends or executes parts of a claims process. Its exposures include unfair outcome, incorrect payment, explanation, appeal, record retention and sector obligations.

Each case uses hypothetical management assumptions to show how the framework operates. The assumptions do not describe a named company or market. A real acquisition requires contracts, production

permission exports, approval records, action traces, incidents, claims, insurance, financial records and legal analysis.

The cases also demonstrate that autonomy is not the sole driver. The contract agent may have lower action volume and high legal consequence. The cloud agent may operate under strong change controls while retaining a large technical blast radius. The claims agent may require a human decision even when evidence preparation is highly automated. Price should follow the verified risk and economic model.

19 Illustrative exposure and remediation economics

The hypothetical treasury case assumes 1.2 million annual actions, a low unauthorised-action probability and a high average consequence, producing expected annual loss and response cost of USD 4.8 million. The contract case assumes 420,000 actions and USD 3.6 million. The cloud case assumes 750,000 actions and USD 7.2 million. The claims case assumes 2.4 million actions and USD 5.4 million. These amounts are management illustrations.

The four cases also assume USD 22 million of one-time remediation across identity, policy enforcement, approval binding, logging, testing, customer work and operating-process change. Recurring incremental control cost is assumed at USD 9 million. The business plan should identify which amount protects existing revenue, enables growth or reduces loss.

A separate probability-weighted liability model assigns USD 29.4 million of expected gross exposure across a payment event, contractual commitment, cloud disruption, data disclosure and regulated-outcome event. It excludes insurance and indemnity recovery because collectibility is uncertain in the illustration. The number is a decision tool, not an accounting estimate.

Sensitivity should vary action volume, event probability, consequence, detection time, recovery, customer concentration, insurance and control effectiveness. Correlated failure deserves a specific case. A shared policy change can expose many customers even when individual workflow history appears strong.

20 Plan the first hundred days around authority continuity

The first objective is continuity of accountable authority. The buyer should preserve production identities, policies, approvals, logs, customer commitments and incident response while establishing ownership. It should identify critical credentials, people, vendors and customer-specific controls before changing platforms.

Integration should proceed by action cohort. Read-only workflows can often move earlier. High-consequence execution should wait for tested identity, policy, approval binding, monitoring and recovery in the buyer environment. Each migration needs a baseline, change plan, test, rollback and customer or regulator decision where required.

The buyer should avoid simultaneous changes to model, prompt, tool, identity and approval. Controlled sequencing helps identify the cause of performance or control change. The target's evidence should remain accessible for earn-out, warranty, customer and regulatory purposes after systems consolidate.

Synergy reporting should include accepted actions, human review, exceptions, incidents, adjusted contribution and cash. Lower labour cost is not a realised synergy if exception cost, customer loss or residual exposure increases. The board should receive a concise authority and liability dashboard tied to the acquisition thesis.

21 Govern agent authority after closing

Post-close governance should assign a business owner, technical owner, security owner, legal or compliance owner and independent assurance role for each material workflow. The operating committee

should approve the action taxonomy, risk appetite, authority limits, customer exceptions and material changes. Consequential decisions should retain accountable human ownership where law, policy or the customer's mandate requires it.

Metrics should include effective permissions, permission surplus, approval volume, rejection and modification, unauthorised attempts, policy failures, reversals, incidents, detection and recovery, customer complaints, service credits, residual exposure and control cost. Measures should be segmented by workflow and customer. A low aggregate incident rate can obscure a concentrated high-consequence cohort.

Change governance should cover models, tools, prompts, policies, identities, thresholds, data and customer configuration. Release evidence should show testing against ordinary, boundary and adversarial scenarios. Critical changes should trigger customer, insurer or regulatory review where required by contract or law.

Independent assurance should test the production system rather than policy alone. Internal audit, external specialists or control functions can sample authority chains, replay decisions and verify remediation. Findings should connect to capital allocation, product scope and customer commitments.

22 Build the investment committee authority record

The investment committee paper should translate technical findings into a decision about value, protection and operating responsibility. It should identify the ten or fewer consequential action paths that drive the largest plausible exposure, the production evidence reviewed for each path, the unresolved gaps and the person accountable for accepting the residual risk. A long catalogue of minor configuration observations can obscure the acquisition issues that change price or the ability to operate.

For each material path, the record should state the principal, delegated purpose, effective identity, permission, enforcement point, human control, action volume, plausible consequence, detection method, recovery route and customer allocation. It should distinguish observed production evidence from management representation and scenario assumption. The committee can then see whether an exposure is already present, contingent on an uncertain event, created by post-close integration or dependent on a planned control improvement.

The paper should reconcile the liability view to the financial model. Control staffing, insurance, customer remediation, technology work and integration delay should appear in forecast cash flow where applicable. Specific claims, provisions or contingent matters should be reviewed with accounting and legal advisers. Price adjustments, escrow, indemnity and contingent consideration should not be treated as substitutes for the operating controls needed to serve customers after closing.

Decision alternatives should be explicit. The buyer may proceed at the proposed price after verification, adjust value, exclude a workflow or entity, delay closing until a condition is satisfied, stage the deployment, require additional seller protection or decline the transaction. Each alternative should state the evidence that would change the decision and the deadline for obtaining it. This structure prevents an unresolved red flag from becoming an unrecorded assumption.

The committee should also approve the post-close authority model. It should name the executive who owns the agentic business, the officers who retain consequential decision authority, the technical and security control owners, the assurance route and the circumstances requiring board escalation. The first-hundred-day budget and sequence should match the approved risk appetite and customer commitments.

Finally, the decision record should preserve dissent and conditions. A reviewer may accept the commercial thesis while requiring a lower authority level until more evidence accumulates. Another may view a customer consent, third-party model dependency or insurance exclusion as a closing issue.

Recording these positions improves accountability and gives the integration team a clear boundary for releasing autonomous capability.

An evidence-release schedule can convert conditions into measurable post-signing work. Each item should specify the baseline record, test population, pass threshold, independent reviewer, failure response and consequence for price, timing or deployment. Evidence should be retained in a controlled repository with stable identifiers so that the committee, transaction counsel, insurers and integration team refer to the same result. When a test fails, the response should address the affected action cohort and customer commitments rather than merely close a ticket. The schedule should expire assumptions that remain unsupported and require renewed approval before authority expands. This discipline connects the investment decision to operating conduct after ownership changes.

Quarterly review should compare residual exposure with the approved acquisition case and document every material departure, owner, remedy and deadline.

Conclusion

An AI-agent acquisition cannot be understood through model capability and software revenue alone. The buyer is acquiring an authority system whose identities, permissions, tools, approvals and actions can create value and liability. The diligence record must therefore prove who can cause what consequence, under whose authority, with which evidence and remedy.

The proposed framework converts that question into transaction work. It defines the perimeter, reconstructs the principal-agent chain, builds the permission graph, classifies consequential action, tests enforcement and human override, quantifies action-cohort exposure and connects findings to valuation and deal protection. It treats current standards and regulations as sources for questions and controls, while requiring target-specific evidence.

The strongest acquisition case is supported by bounded authority, downstream enforcement, meaningful human control, complete action records, aligned customer contracts, tested recovery and a costed post-close operating model. Where evidence remains immature, the buyer can reduce assumed value, require remediation, stage integration or allocate risk through transaction structure. The investment decision should record residual exposure and the person accountable for accepting it.

Authority and permission evidence room

Provide the production principal-agent map, identities, service accounts, groups, roles, tokens, secrets, tool definitions, policy-as-code, permission exports, customer scopes, temporary privilege, break-glass access, revocation tests and change history. Reconcile required and effective permissions and identify permission surplus by consequential action.

Human control and action dossier

Provide action taxonomy, approval rules, reviewer roles, interface captures, approval-to-execution binding, expiry, replay prevention, override and stop tests, review metrics, disagreement, escalation, reversal and recovery evidence. Include representative signed traces from request through final state.

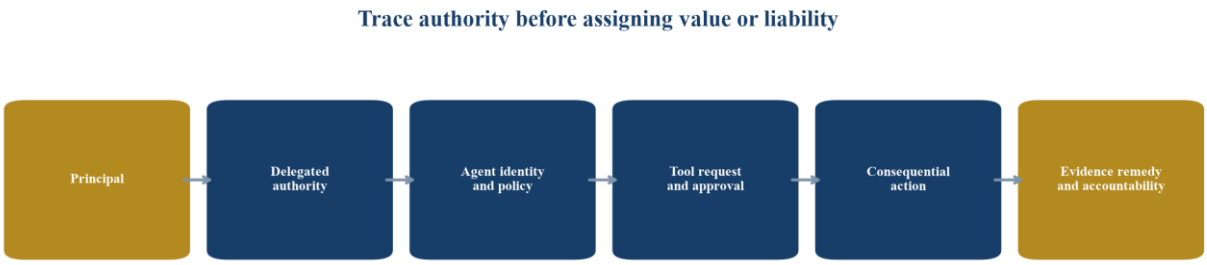
Liability contract and insurance file

Provide customer and provider agreements, negotiated exceptions, complaints, claims, provisions, contingent matters, insurance policies, broker submissions, notices, recoveries and legal analyses. Map technical permissions and operating conduct to contractual authority and liability allocation.

Transaction and integration control file

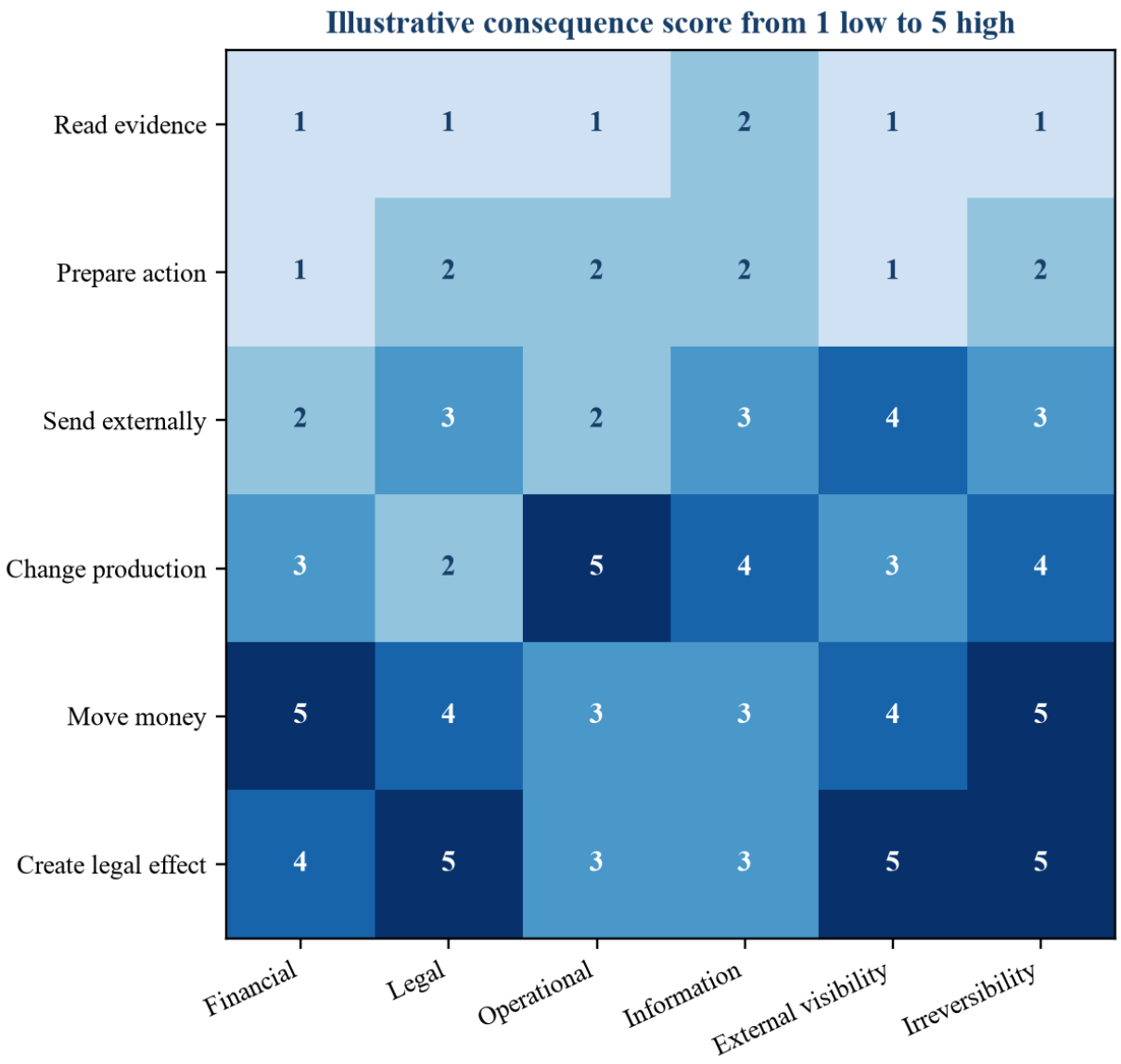
Provide valuation adjustments, exposure scenarios, remediation budgets, representations, disclosures, indemnities, escrow, closing conditions, covenants, contingent consideration, first-hundred-day plan, customer consents, control owners and board reporting. Preserve the evidence needed to test each release of value.

Figure 1 Principal agent authority and consequence chain



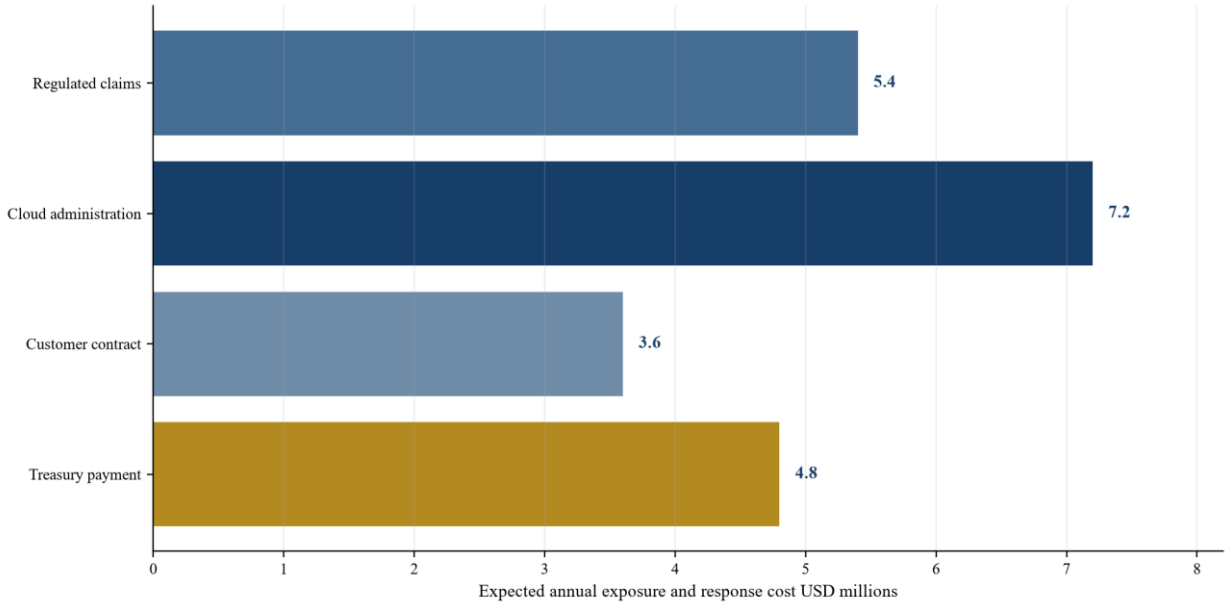
Proposed transaction-diligence chain; every material production action should be attributable and reproducible.

Figure 2 Consequential action control matrix



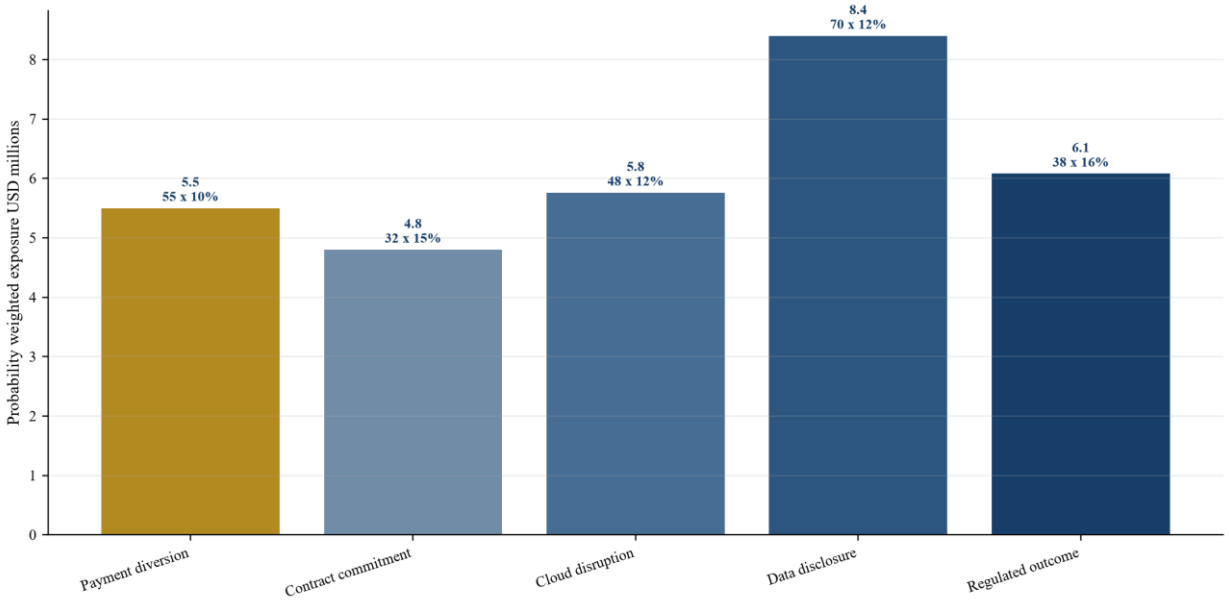
Proposed classification; darker cells require stronger independent approval recovery and transaction protection.

Figure 3 Hypothetical expected annual exposure by agent workflow



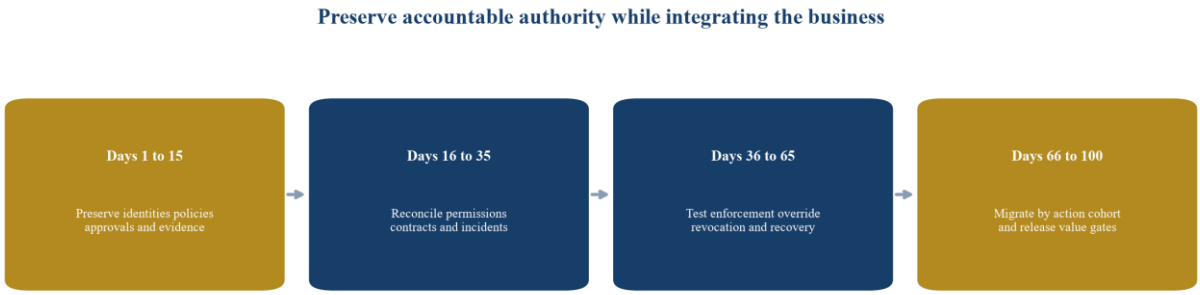
Management assumptions in USD millions; values exclude insurance indemnity recovery tax and purchase-price treatment.

Figure 4 Hypothetical probability weighted gross liability exposure



Management assumptions in USD millions; total expected gross exposure is USD 29.4 million before recoveries.

Figure 5 First hundred days authority continuity sequence



Proposed sequence; timing should follow customer legal regulatory security and integration requirements.

Table 1 Agentic acquisition liability perimeter

Component	Required evidence	Transaction question	Principal failure
Authority	principal purpose delegation and limits	who authorised the action	implied authority without proof
Identity and permissions	production exports roles tokens and policy	what the agent can actually do	permission surplus or shared privilege
Human control	approval record interface timing and independence	can a person prevent or reverse harm	ritual approval or automation bias
Consequence	action cohort volume severity and reversibility	what value or liability can result	technical success hides legal or financial effect
Remedy	detection containment recovery contract and insurance	who bears cost and how quickly	recovery assumed without collectibility

Proposed diligence structure; conclusions require target-specific legal accounting technical and commercial review.

Table 2 Permission and action classification

Level	Agent capability	Human role	Minimum evidence
Observe	discover and read approved evidence	defines purpose and access	source scope identity and audit trace
Recommend	analyse and propose	interprets and decides	rubric sources alternatives and disagreement
Prepare	populate a transaction or change	validates complete action	exact parameters segregation and expiry
Bounded execute	act inside enforceable limits	approves policy and monitors	downstream enforcement signed trace and rollback
Consequential decide	determine rights money safety or regulated outcome	retains accountable authority where required	decision record competence appeal and assurance

Proposed control ladder; workflow-specific law policy and customer requirements govern actual authority.

Table 3 Human override effectiveness test

Dimension	Diligence test	Failure indicator	Transaction response
Information	reviewer sees evidence action and consequence	summary hides parameters or uncertainty	redesign interface and gate release
Authority	reviewer can refuse modify stop and escalate	reviewer lacks role or system permission	reassign authority and enforce segregation
Timing	decision occurs before irreversible execution	approval is retrospective	block execution or add bounded pre-authorisation
Independence	reviewer is separate where consequence requires	agent selects or influences reviewer	independent routing and conflict controls
Workload	review volume permits meaningful attention	near-universal approval or extreme latency	risk-tiering staffing and sampling redesign
Binding	approved action equals executed action	stale altered or replayed approval	cryptographic binding expiry and reconciliation

A visible approval step is insufficient unless every dimension operates in production.

Table 4 Hypothetical annual workflow exposure

Workflow	Annual actions millions	Expected exposure and response cost USD millions	Primary control dependency
Treasury payment agent	1.20	4.8	beneficiary validation value limits and segregation
Customer contract agent	0.42	3.6	approved language apparent authority and version control
Cloud administration agent	0.75	7.2	least privilege change approval and rollback
Regulated claims agent	2.40	5.4	human decision explanation appeal and record retention

Management assumptions; figures are not observations forecasts accounting estimates or valuation conclusions.

Table 5 Hypothetical control remediation budget

Workstream	One time cost	Recurring annual cost	Evidence of completion
Identity and permission redesign	5.0	1.5	production graph least privilege and revocation test
Policy enforcement and approval binding	4.5	1.2	blocked scenarios signed approval and replay rejection
Logging evaluation and incident operations	4.0	2.0	complete trace detection and recovery exercise
Customer contract and configuration remediation	3.5	1.3	agreed authority and technical alignment
Security regulatory and assurance programme	3.0	2.0	tested controls findings and closure
Integration and operating-model change	2.0	1.0	cohort migration acceptance and governance
Total	22.0	9.0	board-approved evidence register

Management assumptions in USD millions; actual scope requires verified architecture customer and legal evidence.

Table 6 Hypothetical probability weighted gross exposure

Event	Gross consequence	Probability	Weighted exposure
Payment diversion	55	10%	5.5
Unauthorised contract commitment	32	15%	4.8
Cloud service disruption	48	12%	5.76
Confidential data disclosure	70	12%	8.4
Incorrect regulated outcome	38	13%	4.94
Total			29.4

Management assumptions in USD millions; expected gross exposure is USD 29.4 million before insurance indemnity recovery tax or purchase-price treatment.

Table 7 Transaction response by evidence state

Evidence state	Finding	Potential transaction response	Post close measure
Verified authority	bounded identities permissions and approvals	support base value and planned integration	permission surplus and policy exceptions
Remediable gap	control weakness with defined fix and cost	price adjustment covenant budget or milestone	closure test and residual exposure
Known exposure	identified event claim or customer exception	disclosure specific indemnity escrow or insurance	claim cash and recovery
Uncertain tail	sparse history or correlated high consequence	scenario discount holdback or restricted deployment	incident leading indicators and assurance
Transfer blocker	missing authority right consent or critical control	closing condition delayed perimeter or no-go	verified transfer and operating acceptance

Proposed framework; actual instruments require current legal tax accounting regulatory insurance and financial advice.

Frequently asked questions

Why are permissions central to an AI agent acquisition

Permissions determine which resources and actions the agent can reach. Model capability becomes a transaction liability when effective authority permits a consequential action. Buyers should reconcile required and effective permissions using production exports and observed execution.

Is a human in the loop enough

The control is effective only when the person receives adequate information, has authority and independence, acts before irreversible execution and can technically stop or change the action. Approval volume, rejection, timing and execution binding should be tested.

How should consequential actions be classified

Classify financial consequence, legal effect, information sensitivity, operational disruption, external visibility, affected population and reversibility. Include chained actions and action velocity. Use the classification to set approval, monitoring, recovery and deal-protection requirements.

What evidence proves delegated authority

Useful evidence links an authenticated principal to a bounded purpose, agent identity, policy, tool, exact action, approval, execution result and retained record. Customer contracts and technical permissions should support the same authority.

How should liability affect valuation

Reconstruct control cost, incident and claims exposure, customer retention, insurance, remediation, integration timing and residual downside. Keep known obligations, contingent matters, remediation and prospective operating risk separate.

Which transaction protections can address agent risk

Depending on verified findings, the parties may use representations, disclosure, specific indemnity, escrow, holdback, insurance, closing conditions, covenants or milestone-based consideration. Legal and accounting advisers must design the actual instrument.

What should happen during the first hundred days

Preserve identities, policies, approvals, logs and customer commitments; reconcile authority; test enforcement, override and recovery; then migrate by action cohort. High-consequence execution should wait for verified controls in the buyer environment.

Which Matchpoint practice is relevant to this work

This research connects to Matchpoint Partners' mergers and acquisitions, technology valuation, transaction diligence, strategy and execution, post-merger integration and capital advisory work for strategic buyers, investors and AI-enabled businesses.

References

[1] IFRS Foundation. IFRS 3 Business Combinations. https://www.ifrs.org/issued-standards/list-of-standards/ifrs-3-business-combinations/

[2] IFRS Foundation. IAS 37 Provisions Contingent Liabilities and Contingent Assets. https://www.ifrs.org/issued-standards/list-of-standards/ias-37-provisions-contingent-liabilities-and-contingent-assets/

[3] IFRS Foundation. Accounting for Contingent Consideration in a Business Combination. https://www.ifrs.org/projects/completed-projects/2013/accounting-for-contingent-consideration-in-a-business-combination/

[4] IFRS Foundation. IAS 38 Intangible Assets. https://www.ifrs.org/issued-standards/list-of-standards/ias-38-intangible-assets/

[5] IFRS Foundation. IFRS 13 Fair Value Measurement. https://www.ifrs.org/issued-standards/list-of-standards/ifrs-13-fair-value-measurement/

[6] International Valuation Standards Council. International Valuation Standards. https://ivsc.org/standards/

[7] National Institute of Standards and Technology. AI Agent Standards Initiative. Updated 14 August 2026. https://www.nist.gov/artificial-intelligence/ai-agent-standards-initiative

[8] National Institute of Standards and Technology. Announcing the AI Agent Standards Initiative. 17 February 2026. https://www.nist.gov/news-events/news/2026/02/announcing-ai-agent-standards-initiative-interoperable-and-secure

[9] National Institute of Standards and Technology. Accelerating the Adoption of Software and AI Agent Identity and Authorization. 2026. https://www.nccoe.nist.gov/projects/accelerating-adoption-software-and-ai-agent-identity-and-authorization

[10] National Institute of Standards and Technology. New Concept Paper on Identity and Authority of Software Agents. 5 February 2026. https://www.nist.gov/news-events/news/2026/02/new-concept-paper-identity-and-authority-software-agents

[11] National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework. https://www.nist.gov/itl/ai-risk-management-framework

[12] National Institute of Standards and Technology. Generative AI Profile NIST AI 600-1. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf

[13] National Institute of Standards and Technology. Cybersecurity Framework 2.0. https://www.nist.gov/cyberframework

[14] National Institute of Standards and Technology. Digital Identity Guidelines. https://pages.nist.gov/800-63-4/

[15] National Institute of Standards and Technology. Zero Trust Architecture SP 800-207. https://csrc.nist.gov/pubs/sp/800/207/final

[16] Cybersecurity and Infrastructure Security Agency. Secure by Design. https://www.cisa.gov/securebydesign

[17] OWASP Foundation. LLM06 2025 Excessive Agency. https://genai.owasp.org/llmrisk/llm062025-excessive-agency/

[18] OWASP Foundation. AI Agent Security Cheat Sheet. https://cheatsheetseries.owasp.org/cheatsheets/AI_Agent_Security_Cheat_Sheet.html

[19] OWASP Foundation. Agentic AI Threats and Mitigations. https://genai.owasp.org/

[20] OWASP Foundation. Agentic AI Security Verification Standard. https://owasp.org/www-project-ai-security-and-privacy-guide/

[21] MITRE. ATLAS Adversarial Threat Landscape for AI Systems. https://atlas.mitre.org/

[22] MITRE. Secure AI Framework. https://www.mitre.org/focus-areas/artificial-intelligence

[23] European Union. Regulation EU 2024 1689 Artificial Intelligence Act Article 14. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

[24] European Commission. AI Act regulatory framework and implementation. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

[25] United Kingdom Information Commissioner's Office. Guidance on AI and data protection. https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/

[26] European Data Protection Board. Automated decision making and profiling guidance. https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines_en

[27] International Organization for Standardization. ISO IEC 42001 Artificial Intelligence Management Systems. https://www.iso.org/standard/42001

[28] International Organization for Standardization. ISO IEC 23894 Artificial Intelligence Risk Management. https://www.iso.org/standard/77304.html

[29] International Organization for Standardization. ISO IEC 27001 Information Security Management Systems. https://www.iso.org/standard/27001

[30] International Organization for Standardization. ISO IEC 27005 Information Security Risk Management. https://www.iso.org/standard/80585.html

[31] OECD. OECD AI Principles. https://oecd.ai/en/ai-principles

[32] OECD. Framework for the Classification of AI Systems. https://oecd.ai/en/classification

[33] World Intellectual Property Organization. Artificial Intelligence and Intellectual Property. https://www.wipo.int/en/web/frontier-technologies/artificial-intelligence/index

[34] United States Federal Trade Commission. Keep your AI claims in check. https://www.ftc.gov/business-guidance/blog/2023/02/keep-your-ai-claims-check

[35] United States Securities and Exchange Commission. Artificial intelligence and investment fraud investor alert. https://www.investor.gov/introduction-investing/general-resources/news-alerts/alerts-bulletins/investor-alerts/artificial-intelligence-fraud

[36] OpenAI. Agents SDK tracing. https://openai.github.io/openai-agents-python/tracing/

[37] OpenAI. Evaluate agent workflows. https://developers.openai.com/api/docs/guides/agent-evals

[38] OpenAI. Trace grading for agents. https://developers.openai.com/api/docs/guides/trace-grading

[39] Anthropic. Responsible Scaling Policy. https://www.anthropic.com/responsible-scaling-policy

[40] Google Cloud. Design patterns for agentic AI systems. https://cloud.google.com/architecture/choose-design-pattern-agentic-ai-system

[41] Microsoft. Agentic AI architecture and responsible AI. https://learn.microsoft.com/en-us/azure/architecture/ai-ml/guide/agentic-ai/agentic-ai-overview

[42] Linux Foundation. Agentic AI Foundation. https://www.linuxfoundation.org/press/linux-foundation-launches-new-agentic-ai-foundation

[43] Cloud Security Alliance. AI Controls Matrix. https://cloudsecurityalliance.org/research/ai-controls-matrix

[44] Center for Internet Security. CIS Critical Security Controls. https://www.cisecurity.org/controls

[45] Internet Engineering Task Force. OAuth 2.0 Security Best Current Practice. https://datatracker.ietf.org/doc/html/rfc9700

[46] Internet Engineering Task Force. JSON Web Token Best Current Practices. https://datatracker.ietf.org/doc/html/rfc8725

[47] OpenID Foundation. Financial grade API Security Profile. https://openid.net/specs/fapi-security-profile-2_0-final.html

[48] National Institute of Standards and Technology. Back to the Future Why Agentic AI Needs a Strong Identity Foundation. 27 August 2026. https://www.nist.gov/blogs/cybersecurity-insights/back-future-why-agentic-ai-needs-strong-identity-foundation

[49] Model Evaluation and Threat Research. Task Completion Time Horizons of Frontier AI Models. https://metr.org/time-horizons/

[50] MLCommons. AI Safety Benchmark. https://mlcommons.org/benchmarks/ai-safety/

About the Author

Chennakeshav (CK) is a corporate finance and investment banking executive with 25+ years of global experience in deal origination, structuring and execution across M&A, growth capital and corporate strategy. He has led value-creation mandates for founders, corporates and funds — bridging the boardroom view to hands-on execution and close.

His career spans Morgan Stanley, HSBC, Lloyds Banking Group, EWEC, ADQ portfolio companies and Emirates Growth Fund, across TMT, real estate, fintech, deeptech, cleantech, infrastructure and energy. He has partnered with C-suite leaders, private equity and venture funds, sovereign wealth funds and family offices to finance complex fund raises and scale-up ventures, and has led M&A due diligence, post-merger integration and business-transformation initiatives to create value.

At Matchpoint Partners he is Managing Partner, leading the firm's corporate finance, M&A and capital-raising practice. He holds an MBA from London Business School, an engineering degree from VTU and a Master of Laws (LLM, in progress) from UCL London.

An active start-up mentor, CK mentors at Techstars, DIFC FinTech Hive, Startup Grind, Founder Institute and IN5, serves as Entrepreneur Mentor in Residence (EMiR) at London Business School, and judges the Entrepreneurship World Cup.

https://www.linkedin.com/in/ckadya/

https://www.matchpoint-partners.com/team/ck-adya.html

This paper is part of a continuing series on the structure of private and alternative markets. The views expressed are the author's own. The paper is for information only, describes market structure in general terms, and does not constitute investment, legal, tax or regulatory advice or a recommendation in respect of any security, vehicle or counterparty.