

Proof before Scale: Governing AI Benefits across a Sponsor Portfolio

An Evidence-Led Operating Model for Attributable Value, Controlled Replication and Board Accountability

Authored by Chennakeshav Adya, Independent Researcher

August 2026

Abstract

Artificial intelligence can improve work quality, cycle time, revenue conversion, risk detection and decision support. A sponsor portfolio, however, creates a measurement problem before it creates a scaling opportunity. Portfolio companies have different data environments, operating processes, control requirements, customer propositions and implementation capacity. A pilot can demonstrate technical feasibility while leaving commercial value, attribution, adoption, total cost and risk unresolved. Scaling weak evidence across many companies can multiply cost and control exposure faster than benefit.

This paper develops a proof-before-scale operating model for sponsors, portfolio-company boards and management teams. It connects the investment thesis, baseline, counterfactual, use-case inventory, causal evidence, full cost, adoption, stage gates, decision rights, jurisdictional controls, value reporting and exit evidence. Six original figures and six tables translate these subjects into a value waterfall, evidence ladder, portfolio heat map, decision matrix, governance architecture and twelve-month operating cadence. The framework draws on current public material from NIST, the European Commission, the UK Government, the Financial Conduct Authority, regulators and standards bodies in the United Arab Emirates, Singapore, Canada, Australia and the United States, together with OECD, ISO, accounting and internal-audit guidance.

All thresholds, timetables, responsibility allocations, scorecards and financial examples in this paper are hypothetical management assumptions for process design. They are not portfolio facts, forecasts or promised outcomes. Each company should adapt the model to its governing law, contractual commitments, regulatory perimeter, accounting policy, risk appetite and operating context. This paper is based on authoritative sources available in August 2026. It provides strategic and operational research and does not replace legal, regulatory, tax, accounting, cyber-security, employment, valuation or investment advice.

JEL Classification: G24, G32, G34, M15, O31, O32, O33

Keywords: artificial intelligence, private equity, portfolio operations, value creation, impact evaluation, AI governance, benefit realisation, causal inference, operating model, sponsor portfolio

1. Make proof an investment discipline

Sponsors are presented with an expanding catalogue of artificial-intelligence opportunities. Some can automate repetitive work. Others can support pricing, forecasting, sales, engineering, customer service, procurement, fraud detection, maintenance or knowledge retrieval. The portfolio advantage appears intuitive: test once, reuse the pattern and spread the benefit. The investment case becomes defensible only when technical performance, operating adoption, attributable economics and control effectiveness are measured together.

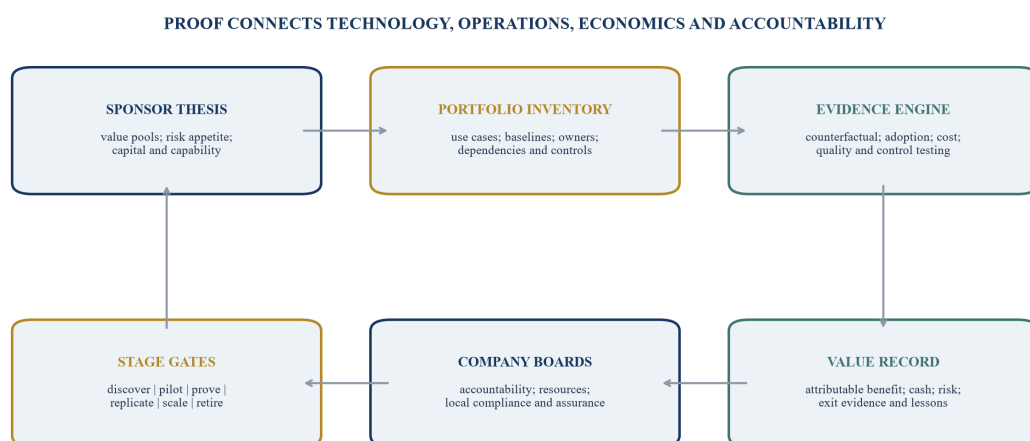
A demonstration proves that a model can produce an output. A pilot tests whether a defined intervention works in a limited operating environment. A proof establishes whether the intervention changes a relevant outcome against a credible baseline and counterfactual. A scale decision asks whether the economics and controls remain acceptable when the user base, data volume, business units, jurisdictions and dependencies expand. These are different decisions and should have different evidence requirements.

The sponsor should therefore govern AI as a portfolio of capital-allocation propositions. Every use case competes for management attention, implementation capacity, data-engineering effort, change resources and control ownership. The decision record should state the value mechanism, affected process, accountable executive, baseline, measurement design, expected adoption, full cost, material risks, stop conditions and evidence needed for the next gate. The record should remain available to the portfolio-company board and relevant sponsor committees.

NIST describes its AI Risk Management Framework as a voluntary resource for organisations designing, developing, deploying or using AI systems [1]. The companion playbook organises actions around Govern, Map, Measure and Manage [2]. That sequence supports an investment discipline: establish accountability, understand the use context, measure performance and impact, then act on the evidence. A sponsor can apply the same structure without forcing every company into the same technology stack.

Proof before scale is also a governance response to enthusiasm risk. A benefit claim becomes more reliable when the team can identify the population affected, the period measured, the change introduced, the control group or comparison, the costs included and the alternative explanations considered. The discipline supports faster replication of strong patterns because the organisation knows what must remain true.

Figure 1. Sponsor portfolio AI benefit-governance system



Scale follows evidence that survives finance, risk, operating and board challenge

The architecture is illustrative. Each sponsor and portfolio company should assign decision rights under its own governance documents and regulatory obligations.

2. Translate the sponsor thesis into testable value mechanisms

An AI thesis should begin with the economic mechanism rather than the model category. A revenue proposition may improve lead qualification, conversion, retention, cross-selling or price realisation. A cost proposition may reduce handling time, error correction, external spend, inventory loss or unplanned downtime. A cash proposition may improve collection, forecast accuracy, working-capital decisions or capital utilisation. A risk proposition may detect anomalies, strengthen surveillance or reduce the expected severity of incidents. A strategic proposition may accelerate product development, increase organisational learning or create a reusable data asset.

Each mechanism needs a chain from intervention to outcome. For example, a service assistant might retrieve approved knowledge faster, which may reduce average handling time, increase first-contact resolution and improve capacity. The financial result depends on actual use, workflow redesign, demand, staffing choices and service quality. Time saved has no automatic cash value. It becomes a financial benefit when the company avoids planned hiring, redeploys capacity into productive work, reduces overtime, increases throughput or improves revenue and retention in a measurable way.

The sponsor should require a value hypothesis with five elements. First, define the operating decision or task that changes. Second, identify the measurable intermediate outcome. Third, state the financial or risk consequence and its timing. Fourth, identify assumptions that management must validate. Fifth, name the executive who can realise the benefit through process, workforce or commercial action. This makes the value case falsifiable.

The hypothesis should also identify possible adverse pathways. Faster content generation can increase review demand. Automated screening can create unfair outcomes or obscure accountability. A coding assistant can increase throughput while introducing security or licensing concerns. A pricing tool can improve margin and harm customer trust if poorly governed. Measuring benefits and harms within the same specification protects the board from a one-sided business case.

The value mechanism becomes a reusable portfolio object when its definition is stable. Sponsors can compare companies on common economic categories while preserving local baselines and implementation details. This supports learning without treating one company's result as proof for another company.

Table 1. AI benefit-measurement specification

Value pathway	Operating measure	Outcome measure	Attribution evidence	Cost and risk adjustment	Benefit owner
revenue conversion	qualified opportunities and response time	incremental gross profit and retention	randomised or phased comparison where feasible	discounts, service cost, churn and model error	chief commercial officer
workforce capacity	minutes per completed task and rework	avoided cost, redeployed output or throughput	task-level logs, time study and matched comparison	licence, integration, review and change cost	operating executive and CFO
working capital	forecast error, dispute cycle and collection action	cash conversion and financing cost	cohort or interrupted-time-series analysis	seasonality, policy changes and customer mix	CFO and treasury lead
asset reliability	warning precision, downtime and maintenance action	avoided downtime and lifecycle cost	controlled rollout or matched-asset analysis	false alarms, sensors, maintenance and disruption	operations director
risk detection	alert quality, investigation time and control coverage	expected-loss reduction and control effectiveness	back-testing, parallel run and incident analysis	missed events, bias, privacy and assurance cost	chief risk officer
product development	experiment cycle and release quality	time to revenue, adoption and contribution margin	milestone comparison and customer evidence	technical debt, IP, safety and support cost	product executive

The entries are an illustrative management template. Metrics, thresholds and accounting treatment require company-specific approval.

3. Establish the baseline before the intervention

A baseline describes the process and outcome before the AI intervention. It should cover enough time to represent normal variation and should segment the population where performance differs materially. A single portfolio-wide average can hide differences by company, product, geography, customer, channel, employee group or asset type. The baseline period should record seasonality, prior initiatives, staffing changes, price changes, acquisitions, system migrations and external events that may affect the outcome.

The UK Government's 2026 Green Book places appraisal within a structured process for developing and evaluating interventions [12]. The Magenta Book provides guidance on evaluation, and its guidance on AI interventions highlights the need to understand whether an intervention caused the observed outcomes [13, 14]. A sponsor can apply these principles to a commercial portfolio: define the problem, articulate the theory of change, select an appropriate evaluation design and collect evidence before the decision window closes.

Baseline quality should be scored before a pilot launches. The team should assess data definition, completeness, timeliness, lineage, stability and representativeness. It should confirm whether people record the process consistently and whether measures can be independently reproduced. Weak baseline data does not always prevent a pilot. It changes the claim that the pilot can support and may require a preliminary measurement phase.

The baseline should include cost. Existing process cost includes direct labour, external services, systems, rework, delay, incidents and management oversight. The AI intervention adds licence or compute, integration, data preparation, security, monitoring, human review, training, vendor management, legal analysis and model-change cost. Comparing a fully loaded intervention cost

with a partially measured legacy cost can bias the decision in either direction.

Management should approve a measurement freeze that records definitions before results are observed. Definitions may need to change when the pilot reveals a flaw. Every change should be dated, explained and reflected in the analysis. This practice reduces the temptation to select a favourable metric after the fact.

4. Choose a credible counterfactual

The counterfactual estimates what would have happened without the intervention. It is the foundation of an attributable benefit claim. A before-and-after comparison can be informative, although it may confuse the effect of AI with seasonality, demand, management attention, process redesign, training or another initiative. Stronger designs create a more credible comparison.

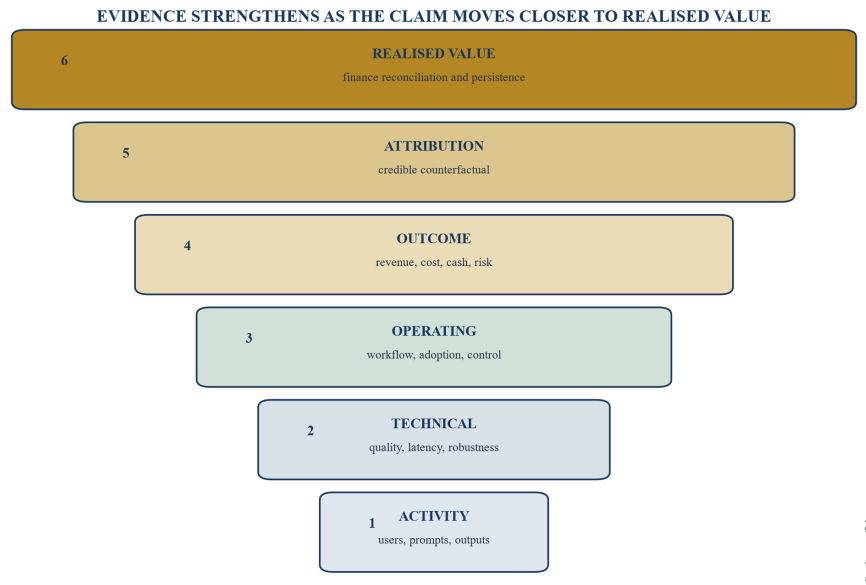
Randomised assignment is useful when eligible users, transactions, cases or sites can be allocated fairly and operationally. A phased rollout can compare early and later groups while eventually giving all eligible groups access. A matched comparison can pair treated units with similar untreated units. Difference-in-differences can compare changes across treated and comparison groups when assumptions are plausible. An interrupted time series can test whether the level or trend changed after deployment. A regression discontinuity may work where a clear threshold determines treatment. Qualitative evidence remains important for understanding mechanisms, adoption and unintended effects.

The UK Test and Learn guide explains experimentation approaches for policy and service interventions [15]. The same logic can be used in a company with proportionate controls. The ethical and legal context still matters. Randomisation should not deny required protections, create discrimination or conflict with employment and customer obligations. The design should be reviewed by the relevant risk, legal and operating owners.

Some use cases cannot support a strong experimental design. A rare fraud event, major transaction, safety incident or one-off strategic decision may have no meaningful control group. The team should then triangulate evidence: back-testing, simulation, parallel operation, expert review, process metrics, historical cases and sensitivity analysis. The claim should reflect the strength of the evidence.

A sponsor should maintain an evidence ladder. Technical accuracy and user satisfaction are early evidence. Reliable process change, sustained adoption and controlled operation are stronger. Attributable P&L, cash or risk outcomes with full-cost reconciliation are stronger still. A scale decision can proceed at different levels depending on reversibility, downside, materiality and control readiness.

Figure 2. Evidence ladder from activity to attributable portfolio value



Advancement is illustrative. The approving body should set the evidence standard according to materiality, reversibility and risk.

Table 2. Evaluation-design selection guide

Operating situation	Potential design	What it can support	Principal limitation	Minimum control
many comparable cases or users	randomised controlled test	strong estimate of average causal effect	contamination, fairness and local variation	pre-defined outcome and allocation protocol
rollout must reach every unit	stepped or phased rollout	comparison across rollout waves	timing may correlate with readiness	documented sequencing and common measurement
comparable untreated units exist	matched comparison or difference-in-differences	effect estimate under stated assumptions	unobserved differences and parallel-trend risk	matching logic and pre-trend test
long, stable measurement history	interrupted time series	change in level or trend	concurrent events can confound result	sufficient periods and event register
threshold determines eligibility	regression discontinuity	local effect around threshold	limited generalisability	stable rule and no manipulation around cut-off
rare or high-consequence event	back-test, simulation and parallel run	performance and control evidence	limited evidence of live financial effect	scenario library, expert challenge and stop rules

The guide is illustrative. Statistical, legal, ethical and operational specialists should validate the selected design.

5. Build a complete use-case inventory

A sponsor cannot govern what it has not inventoried. Portfolio companies may use enterprise tools with embedded AI, specialist applications, vendor APIs, internally developed models, employee-accessed public services and AI features added through software updates. Procurement records alone will miss experimentation and embedded functionality. An inventory should combine technology, procurement, privacy, cyber-security, legal, risk and business-owner inputs.

Each record should identify the company, business process, decision supported, users, affected people, data categories, model or service, vendor, deployment method, human oversight, material dependencies, jurisdiction, intended benefit, accountable executive and lifecycle status. It should also identify whether the output informs an employment, credit, insurance, health, safety, pricing, customer-access or other consequential decision.

The inventory should distinguish an assistant from an autonomous action. It should record who reviews the output, who can override it, what happens when the system fails and whether the decision can be reconstructed. High-volume low-value assistance may create aggregate operational exposure. A narrow model used in a material decision may require deeper validation even with few users.

Canada's Algorithmic Impact Assessment is a questionnaire designed to assess the impact level of automated decision systems under the federal Directive on Automated Decision-Making [24, 25]. Australia's Guidance for AI Adoption Practices asks organisations to establish accountability, assess impacts, manage data, test performance and maintain human oversight [26]. These are public-sector or national guidance tools with their own scope. Their structured questions can inform a commercial inventory without being treated as universally binding requirements.

The inventory should be a live control object rather than a survey archive. New uses enter through intake. Material changes trigger reassessment. Retired systems retain an evidence trail covering data, decisions, incidents and obligations. Portfolio reporting should aggregate consistent fields while allowing local annexes for regulated or sector-specific requirements.

6. Tier risk and value separately

Value potential and risk are different dimensions. A high-value use case can merit investment and intensive controls. A low-value use case with high risk may deserve redesign or retirement. A low-risk, reversible productivity tool may advance with lighter evidence. Combining value and risk into one score can conceal the trade-off.

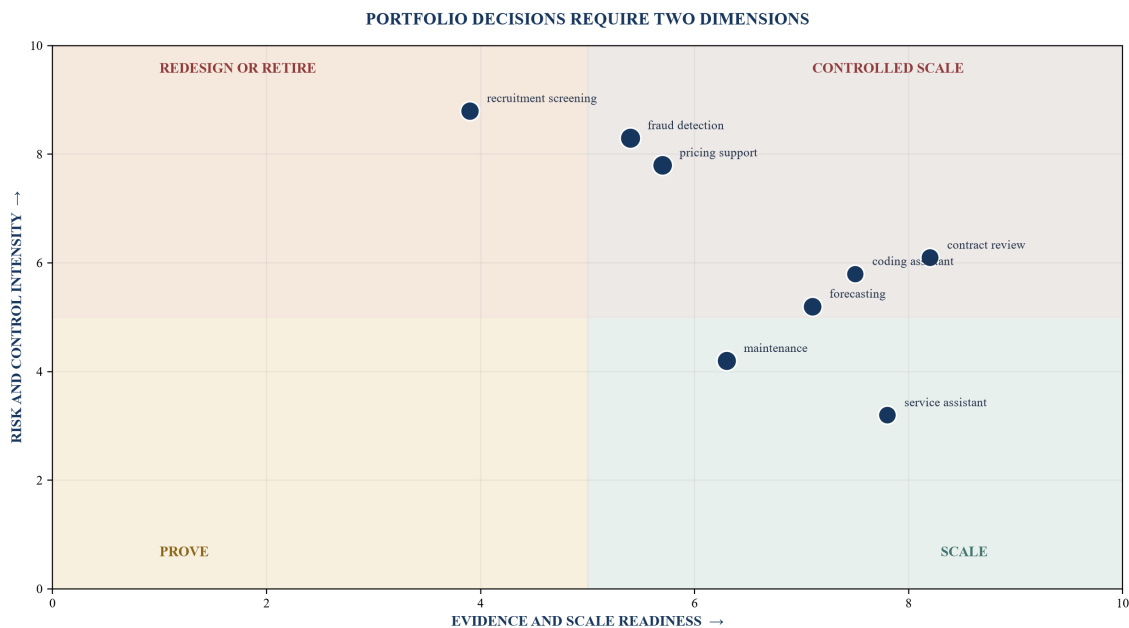
Value potential should consider addressable economic pool, operational reach, time to benefit, persistence, cash conversion, strategic relevance and replication potential. Evidence readiness should consider baseline quality, outcome measurability, comparison feasibility, sample size, implementation stability and finance traceability. Risk should consider affected rights, financial materiality, safety, privacy, security, bias, explainability, customer impact, model dependence, vendor concentration and reversibility. Scale readiness should consider process ownership, data, integration, training, support, monitoring and incident response.

The EU AI Act uses a risk-based structure and sets requirements for defined high-risk systems [3, 4]. The official Commission timeline reflects amendments enacted through the 2026 AI Omnibus; application dates vary by provision and category [4]. A portfolio inventory should therefore map the actual provider, deployer, importer and distributor roles, prohibited-practice analysis, system category, territory and implementation date. A generic label of high, medium or low does not replace legal classification.

The portfolio heat map should show evidence strength alongside risk and value. The sponsor can use it to decide where a shared capability is useful. Shared procurement, model inventory, cyber testing, legal analysis, measurement templates and vendor due diligence can reduce duplication.

Local boards should retain decisions that depend on their workforce, customers, regulated activities and company-specific risk appetite.

Figure 3. Portfolio heat map for proof and scale decisions



Positions are illustrative management examples. Actual use cases require company-specific evidence and risk assessment.

7. Measure adoption as part of the intervention

AI benefit depends on how people use the system and how the surrounding process changes. A technically accurate model may produce little value when users do not trust it, managers do not change decisions, exceptions overwhelm the workflow or incentives reward the old process. Adoption should therefore sit inside the causal model rather than appear as a communication metric after deployment.

The measurement plan should distinguish access, activation, frequency, eligible-task coverage, appropriate use, override, escalation, rework and sustained use. High usage can be a warning when employees use a tool outside approved cases or rely on it without required review. Low usage can reflect poor design, weak training, insufficient data, slow response, unclear accountability or a sensible rejection by experienced users.

Management should observe the work. Interviews, task analysis and process mining can reveal where AI adds steps, shifts burden or creates hidden review work. Front-line staff should be able to report failure modes without being measured only on adoption. Customers and affected people may provide evidence about quality, clarity, accessibility and trust.

The operating model should specify the decision that remains human. Human oversight needs authority, competence, time and information. A nominal approval click offers limited protection when the reviewer cannot understand the basis of the output, faces unrealistic volumes or is penalised for disagreement. Override rates should be analysed by context and outcome. Very high or very low rates can both require investigation.

Workforce impact should include task redesign, skill development, workload, role clarity and employee relations. Capacity claims should identify what management will do with released time. The plan may redeploy people, absorb growth, improve service, reduce external spend or avoid

vacancies. Each route has a different financial timing and human consequence.

8. Reconcile full cost and realised benefit

A gross benefit estimate should be reduced through a consistent value waterfall. Start with the addressable value pool and the effect observed in the eligible population. Apply the attributable share supported by the evaluation. Adjust for adoption, persistence, leakage and overlap with other initiatives. Deduct one-off and recurring costs. Translate operating gains into recognised P&L, cash, risk or strategic outcomes under the company's policies.

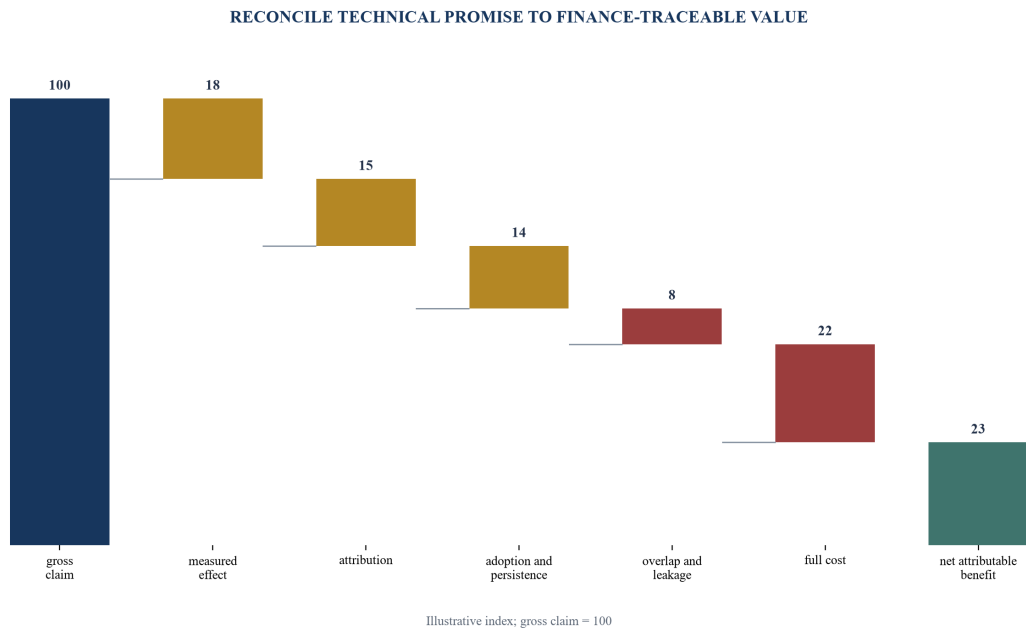
Full cost includes more than a software subscription. It can include data preparation, architecture, integration, model access, compute, licences, vendor services, internal engineering, cyber-security, privacy, legal review, assurance, training, process redesign, support, monitoring, incident response, decommissioning and management time. It should also include the opportunity cost of scarce teams and the cost of maintaining parallel processes during transition.

Benefits should be classified to avoid false precision. Finance-verified realised benefits have entered the accounts or cash flow under an agreed bridge. Committed benefits arise from approved actions, such as removal of a vendor cost or avoidance of a contracted hire. Validated operating benefits have credible outcome evidence but have not yet translated into finance. Strategic option value may be real and should remain separate from booked benefit.

IAS 38 addresses recognition and measurement of intangible assets and distinguishes research from development expenditure under specified conditions [30]. AI investment may involve software, data, licences, services and internal development with different accounting treatments. The finance team should establish the applicable policy before the board sees a return calculation. An economic benefit model and an accounting treatment answer different questions.

The sponsor should prohibit double counting across use cases. A revenue increase cannot be attributed in full to a pricing model, sales assistant and marketing system when they affect the same customer outcome. A central benefit register should assign primary and contributory mechanisms and reconcile to company results.

Figure 4. AI value waterfall from claim to attributable net benefit



Values are hypothetical management assumptions used to illustrate the reconciliation method. They are not forecasts or portfolio results.

9. Define a proof packet before the pilot starts

Every pilot should produce a standard proof packet. The packet should be designed before launch and updated through the lifecycle. It becomes the evidence base for the scale decision, portfolio learning and future diligence.

The packet should contain the approved problem statement, value hypothesis, process map, baseline, eligible population, evaluation design, data dictionary, risk and impact assessment, architecture, vendor diligence, security assessment, privacy analysis, model testing, human-oversight design, training plan, incident procedure, cost model, adoption plan, benefits bridge, decision log and lessons. It should identify gaps and unresolved assumptions.

Evidence should be reproducible. The analysis should preserve the code or calculation logic, dataset versions, sampling rules, exclusions, model version, prompt or configuration where relevant, review protocol and approval date. Access controls should protect personal, confidential and proprietary information. The retention schedule should reflect legal, regulatory and contractual obligations.

The finance controller should challenge the economic bridge. Risk and compliance should challenge use classification, controls and incidents. Technology and security should challenge architecture and dependencies. The operating executive should confirm whether the process change is sustainable. Internal audit or an independent assurance provider may review the governance design or selected high-materiality use cases.

The Institute of Internal Auditors' updated AI Auditing Framework provides guidance on AI governance, management and internal-audit activity [21]. ISO/IEC 42001 specifies requirements for an AI management system, while ISO/IEC 42005:2025 addresses AI system impact assessment and ISO/IEC 42006:2025 addresses bodies providing audit and certification of AI management systems [19, 20]. Certification is a separate decision. The underlying

management-system disciplines can help structure evidence and accountability.

10. Use stage gates that can stop, narrow or redesign work

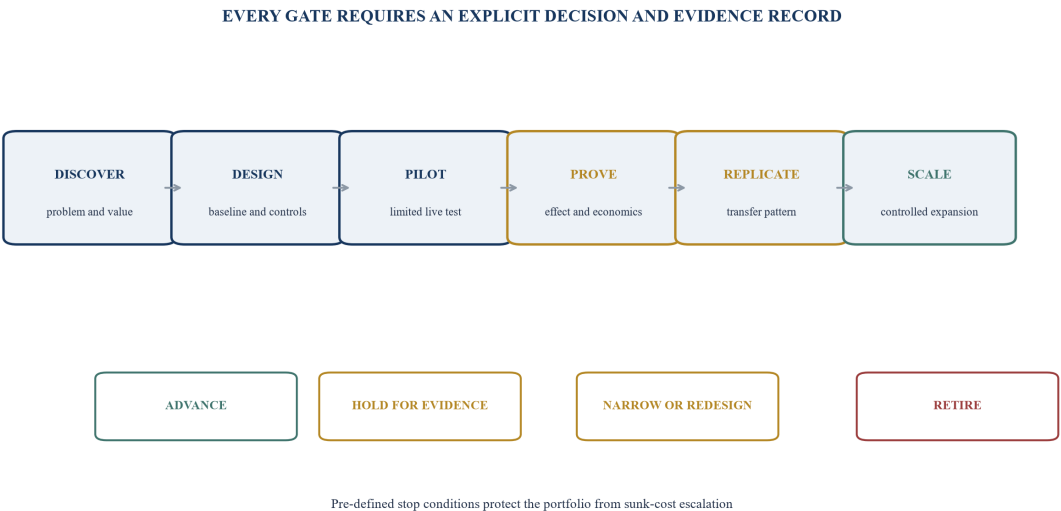
A stage gate should be a decision, not a presentation milestone. The approving body should be able to advance, hold, narrow, redesign, return for evidence or retire the use case. The submission should state what changed since the prior gate and which conditions remain open.

The discovery gate confirms that the problem is material, the value mechanism is plausible and an AI intervention is an appropriate candidate. The design gate approves the baseline, evaluation, data, architecture, risk classification and resources. The pilot gate authorises limited live use under specified controls. The proof gate assesses outcome, attribution, adoption, cost and incidents. The replication gate tests whether the operating pattern transfers to another team or company. The scale gate commits wider resources and control capacity. The retirement gate closes the system and preserves required records.

Gate criteria should vary with reversibility and consequence. A low-risk assistant can be rolled back quickly and may advance with a smaller sample. A system affecting employment, customer access, financial decisions, safety or regulated activity may require stronger testing, independent challenge and staged exposure. Scale can be conditional: a company may extend to a defined population while holding features or jurisdictions outside the approved perimeter.

Stop conditions should be set in advance. They may include unacceptable error, bias, security weakness, data leakage, control failure, customer harm, poor adoption, unstable cost, vendor breach or failure to achieve the minimum outcome. A stop condition protects decision quality when sunk cost and executive sponsorship create pressure to continue.

Figure 5. Proof-before-scale stage-gate path



Gate labels and periods are illustrative. Companies should set criteria according to use-case materiality and control requirements.

Table 3. Stage-gate decision scorecard

Gate	Required question	Core evidence	Typical decision owner	Possible decision
discover	is the problem worth solving and is AI suitable?	value pool, process map, alternatives and accountable owner	company executive committee	stop, explore or design
design	can the pilot generate decision-grade evidence safely?	baseline, counterfactual, risk assessment, architecture and budget	company board delegate	approve, narrow or redesign
pilot	is limited live operation controlled?	validation, user readiness, incident path and data controls	operating and risk owners	launch, hold or stop
prove	is the result attributable, adopted and economical?	outcome analysis, full cost, adoption, incidents and finance bridge	board or investment committee	replicate, extend or retire
replicate	does the pattern transfer beyond the first setting?	second-site evidence, local controls and transfer cost	sponsor and receiving company	scale, localise or stop transfer
scale	can governance and operations sustain expanded use?	capacity, monitoring, assurance, vendor resilience and benefit plan	company board with sponsor oversight	scale conditionally or defer

The thresholds are an illustrative management design and require sponsor and company approval.

11. Allocate sponsor and portfolio-company decision rights

The sponsor can create leverage through standards, specialist capability, procurement and cross-portfolio learning. Each portfolio company remains a separate legal and operating entity with its own board, executives, employees, customers, contracts, regulators and data. Decision rights should respect that boundary.

The sponsor should own the portfolio thesis, minimum evidence standard, aggregate reporting, shared capability, cross-company pattern library and capital allocation across sponsor-funded initiatives. It may coordinate vendor diligence, architecture principles, security testing, regulatory monitoring and expert support. It should identify concentration where many companies depend on the same provider, model, cloud platform or data service.

The company board should approve material use within its business, risk appetite, resources and obligations. Management should own process redesign, data quality, user training, customer and employee impacts, control operation, benefit delivery and incident response. Finance should reconcile value. Risk, compliance, privacy, security and legal functions should exercise their established authorities.

A portfolio AI council can coordinate without becoming a shadow board. Its charter should define advisory and approval responsibilities. Meeting records should identify company-specific decisions and escalation. Sponsors should avoid performance reporting that pressures management to suppress adverse evidence or continue a weak pilot.

Decision rights should also address intellectual property, data sharing and reusable components. A pattern learned at one company may involve confidential information, customer commitments or employee data. Cross-portfolio transfer requires a documented legal and commercial basis. Shared code or prompts need ownership, security and maintenance arrangements.

Table 4. Sponsor and portfolio-company decision-rights matrix

Decision	Sponsor role	Company board role	Management role	Independent challenge
portfolio AI thesis	approve value pools and minimum standards	consider relevance to company strategy	identify operating opportunities	investment committee and operating advisers
use-case approval	set portfolio reporting and sponsor-capital conditions	approve material local deployment	design, implement and operate	risk, legal, security and employee representatives as relevant
proof design	provide methods and specialist support	confirm proportionality and accountability	establish baseline, test and document	finance controller, data scientist or evaluator
vendor and platform	identify portfolio concentration and commercial leverage	approve material dependency	procure, integrate and monitor	cyber-security, privacy and legal teams
scale decision	compare evidence and replication economics	authorise company exposure and resources	execute change and realise benefit	assurance or internal audit for selected cases
portfolio reporting	aggregate consistent measures and lessons	verify company submission	maintain evidence and explain variance	finance and risk committees

The matrix is illustrative. Corporate governance documents, shareholder rights and local law determine actual authority.

12. Govern data, models, vendors and cyber dependencies

AI systems depend on data, models, infrastructure, integrations and vendors. The proof must cover the system as operated, not an isolated benchmark. A model can perform well in a test set and fail when the input distribution changes, users alter prompts, retrieval sources become stale, an upstream system changes or a vendor releases a new version.

Data governance should identify source, purpose, legal basis, quality, lineage, access, retention and permitted reuse. Training, fine-tuning, retrieval and logging create different data paths. Confidential and personal data should not enter a service merely because a user can paste it. Controls can include approved interfaces, classification, redaction, retrieval boundaries, role-based access, encryption, logging and deletion procedures.

Model governance should record intended use, limitations, evaluation datasets, performance metrics, robustness, bias testing, explainability needs, human oversight, version and change history. Generative systems may require tests for factual reliability, harmful output, prompt injection, data leakage and insecure tool use. Performance thresholds should reflect the operating consequence of error.

Vendor diligence should address security, privacy, intellectual property, subcontractors, data location, model changes, service levels, audit rights, incident notification, exit, portability, financial resilience and regulatory support. Portfolio concentration should be visible. A common

provider can create purchasing leverage and correlated failure.

NIST's Generative AI Profile describes risks and suggested actions for generative AI [5]. The joint UAE financial-regulator guidelines on enabling technologies address governance, risk management, data, cyber-security and third-party considerations for financial institutions [18]. These sources have defined scopes. They support a general principle for sponsor portfolios: the control perimeter follows the full technology and service chain.

13. Apply jurisdictional and sector controls to the actual use

Portfolio governance needs a common core and local overlays. The common core can cover inventory, accountability, value evidence, impact assessment, data, testing, human oversight, monitoring, incidents and documentation. The overlay maps the company's location, sector, customers, affected people, role in the AI supply chain and contractual commitments.

In the European Union, the AI Act's obligations depend on system classification and actor role [3, 4]. Data protection remains relevant where personal data is processed. The European Data Protection Board's Opinion 28/2024 addresses aspects of AI models under the GDPR, including anonymity, legitimate interests and consequences of unlawful personal-data processing [6].

In the United Kingdom, the FCA states that its principles-based framework and existing rules apply where regulated firms use AI [10]. The AI Management Essentials guidance provides a structured tool for organisations [11]. In the UAE, national AI ethics principles address fairness, accountability, transparency and related safeguards [16]. The CBUAE issued guidance in February 2026 on consumer protection and responsible AI adoption by licensed financial institutions [17]. The DFSA's 2025 survey reported rapid adoption and continuing governance gaps among surveyed DIFC firms [18].

Singapore's Model AI Governance Framework for Generative AI identifies dimensions including accountability, data, trusted development and deployment, incident reporting, testing and security [22]. Australia's Guidance for AI Adoption Practices provides ten practices, while ASIC's review of 23 financial-services licensees identified 624 AI use cases and governance gaps [26, 27]. The US Government Accountability Office framework organises accountability around governance, data, performance and monitoring [8]. US securities regulators have also taken enforcement action against misleading AI statements [9].

Local mapping should be maintained as obligations develop. The sponsor should use qualified advisers and regulator publications for each material deployment. A portfolio standard can set a minimum; it cannot substitute for use-specific legal analysis.

Table 5. Jurisdictional control overlay for a sponsor portfolio

Jurisdiction or framework	Current authoritative anchor	Portfolio question	Evidence to retain
international	OECD AI Principles; ISO/IEC 42001 and 42005	does the management system assign accountability and assess impacts?	policy, impact assessment, objectives, controls and review
European Union	AI Act and GDPR-related guidance	what is the system category and actor role; when do obligations apply?	classification, role map, technical and deployer documentation
United Kingdom	FCA AI approach and AI Management Essentials	which existing regulatory and governance requirements apply?	accountability, testing, customer-impact and operational-resilience evidence
United Arab Emirates	national AI ethics principles; CBUAE and financial-regulator guidance	which federal, financial-centre and sector requirements govern the use?	ethics assessment, customer protection, model and third-party controls
Singapore	Model AI Governance Framework for Generative AI	how are accountability, data, testing, security and incidents governed?	system map, evaluation, security testing and incident record
Canada	Directive on Automated Decision-Making and impact assessment	can structured impact questions improve the control design?	impact score, peer review, notice and monitoring where applicable
Australia	Guidance for AI Adoption Practices; ASIC review	are governance arrangements keeping pace with adoption?	accountable owner, inventory, impact assessment, testing and monitoring
United States	NIST AI RMF; GAO accountability framework; sector rules	which voluntary framework and binding requirements fit the actual use?	governance, data, performance, monitoring and disclosure evidence

This is a research map, not legal advice. Applicability depends on the system, entity, sector, role, location and affected people.

14. Protect external claims and valuation narratives

An AI benefit claim can influence customers, employees, investors, lenders and buyers. External statements should be connected to evidence and reviewed through existing disclosure controls. A company should be able to explain what the system does, where it is used, how the outcome was measured, which costs were included and which risks remain.

The SEC charged two investment advisers in 2024 over false and misleading statements about their use of artificial intelligence [9]. That enforcement concerned specific securities-law facts and firms. The broader governance lesson is useful: labels such as AI-powered or AI-driven should have a documented basis. Marketing, fundraising and transaction materials should avoid implying deployment, capability or performance beyond the evidence.

At exit, a buyer may test whether the claimed benefit is durable and transferable. Diligence can cover licences, data rights, architecture, model dependencies, cyber-security, privacy, regulatory classification, workforce adoption, vendor concentration, cost, incidents and evidence of financial contribution. A pilot with a strong headline and weak documentation can create diligence friction.

The exit evidence room should reconcile operating measures to the financial record. It should identify which benefits have been realised, committed, validated or estimated. It should show whether the result persists after novelty, management attention and temporary support decline. It should separate value attributable to AI from broader transformation and market changes.

Valuation should follow the applicable method and evidence. AI may influence forecast growth, margin, capital intensity, risk, competitive position or option value. It can also add cost, dependency and execution risk. The valuation bridge should expose these channels and avoid applying an unsupported premium merely because the company uses AI.

15. Decide when a shared portfolio platform creates value

Sponsors may consider a shared AI platform, preferred vendor panel, centre of excellence or common data service. The case depends on repeatability, scale economics, control consistency and local fit. A shared platform can reduce duplicate procurement, improve security, create common evaluation tools and accelerate learning. It can also create concentration, complex data boundaries, slow local innovation and lock the portfolio into a vendor or architecture.

The decision should separate common capabilities from company-specific applications. Common components may include identity, secure model access, logging, evaluation libraries, model inventory, vendor diligence, policy templates, training and incident playbooks. Company-specific components may include process integration, proprietary data, domain models, customer-facing design and local regulatory controls.

Transfer evidence should show that the use case mechanism survives a new context. A service assistant proven in one company may depend on a well-maintained knowledge base, stable product set, call volume, user skill and quality process. The receiving company should test these assumptions. Replication is a new decision with prior evidence, not an automatic rollout.

The business case should allocate shared cost transparently. Sponsor-funded development, company charges, usage fees and exit obligations affect incentives. Companies should understand data and intellectual-property rights, support commitments and what happens when ownership changes. A buyer may require continued access, transition services or separation.

Scale economics should include capacity for governance. More users and companies require monitoring, access management, training, support, model-change review and incident response. The platform should not grow faster than the sponsor and companies can control it.

16. Prevent double counting across the portfolio

Portfolio reporting creates a temptation to aggregate every approved business case. The resulting number may double count shared outcomes, ignore cannibalisation and combine different evidence classes. A central value ledger should therefore sit alongside the use-case inventory.

Each benefit record should identify company, use case, value pathway, baseline, counterfactual, measurement period, gross effect, attributable share, adoption adjustment, cost, accounting treatment, cash timing, evidence class, approver and overlap group. Overlap groups link interventions that affect the same outcome. Finance should nominate a primary driver or allocate contribution through an approved method.

The ledger should reconcile company-level claims to the management accounts and cash flow where relevant. Aggregate portfolio reporting should show gross validated benefit, deductions, realised value and forecast or option value separately. It should also show discontinued use cases and negative outcomes. Excluding failed pilots produces a biased view of the programme.

Foreign exchange, acquisitions, disposals and changes in ownership period should be treated consistently. The sponsor should disclose whether a benefit is annualised, cumulative or within the holding period. Run-rate figures should not be added to realised cash without a clear bridge.

The value ledger can also support capital allocation. A use case with moderate benefit and high confidence may deserve replication before a larger but weakly evidenced claim. A failed pilot can still create value when it identifies an invalid assumption early and prevents portfolio-wide spend. The ledger should capture avoided investment as a decision outcome without reporting it as realised earnings.

17. Run a board and investment-committee dashboard

A board dashboard should support decisions rather than celebrate activity. It should show the use-case population by stage, value pathway, evidence class and risk tier. It should identify decisions due, conditions open, material incidents, vendor concentration, benefits reconciled by finance, costs committed and capacity constraints.

The dashboard should separate leading and lagging indicators. Leading indicators include baseline readiness, evaluation quality, user training, control completion, adoption and issue closure. Lagging indicators include operating outcome, attributable financial benefit, incidents, customer impact and persistence. Both are needed. Activity metrics such as number of models or users should have context.

Material use cases should have a one-page decision summary. It should state the hypothesis, evidence, economics, risk, unresolved issues, management recommendation and requested decision. The committee should see sensitivity to adoption, effect size, cost and persistence. It should also see the consequence of waiting, narrowing or stopping.

Portfolio comparisons should avoid false league tables. Companies differ in size, sector, maturity and opportunity. Normalised measures such as benefit relative to addressable cost, time to proof and evidence quality can help. The purpose is to allocate support and learn, not reward aggressive reporting.

The dashboard should link to the proof packet and decision log. Every figure should have an owner, definition and update date. Finance and risk sign-off should be visible where required. The record becomes a governance asset for internal audit, lender diligence, fundraising and exit.

18. Operate a twelve-month sponsor cadence

The first hundred days should establish governance and measurement capacity. The sponsor can appoint accountable leaders, approve the taxonomy, inventory current uses, identify high-value pathways, set minimum evidence standards, select a small number of pilots and establish the value ledger. Portfolio companies should validate local obligations and resources.

The next phase should generate proofs. Teams collect baselines, launch controlled pilots, monitor adoption and risk, reconcile cost and bring decisions to gates. Shared specialists can support

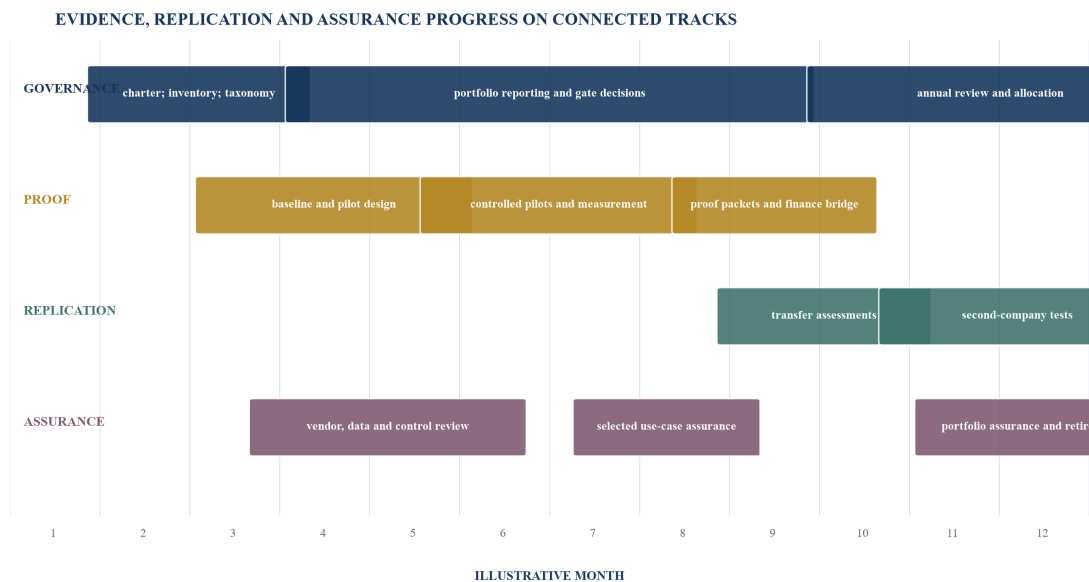
evaluation, security, vendor contracting and change design. The sponsor should publish lessons within confidentiality boundaries.

Replication should begin only after the first proof packet survives challenge. The receiving company performs a transfer assessment, confirms data and process conditions, and establishes its own baseline. A second successful context strengthens the pattern. It does not remove the need for local oversight.

The annual review should assess portfolio value, failed tests, concentration, incidents, regulatory developments, capability and next-year capital allocation. It should retire unused systems, close obsolete access and refresh vendor and model assessments. The board should understand how AI exposure has changed.

The cadence should leave room for urgent opportunities and incidents. A fast-track path can exist for low-risk, reversible experiments with explicit limits. Emergency suspension authority should be clear. Programme governance should accelerate good decisions by making evidence and authority visible.

Figure 6. Twelve-month proof-before-scale operating cadence



Timing is an illustrative management assumption. Actual sequencing depends on portfolio readiness, use-case risk and resources.

Table 6. Board dashboard and operating cadence

Dashboard element	Decision supported	Primary owner	Illustrative frequency	Evidence source
use-case inventory by stage and risk	attention, capacity and escalation	portfolio AI lead	monthly	inventory and decision log
proof quality and open assumptions	pilot continuation and scale readiness	evaluation lead	at each gate	baseline, design and proof packet
attributable value and full cost	capital allocation and benefit recognition	company CFO and sponsor finance	monthly or quarterly	value ledger and management accounts
adoption and process change	intervention redesign and workforce action	operating executive	monthly during pilot	workflow logs, surveys and observation
incidents and control effectiveness	suspension, remediation and assurance	risk and security leaders	immediate plus monthly summary	incident, monitoring and control records
vendor and platform concentration	resilience, negotiation and exit planning	technology and procurement leads	quarterly	architecture and contract inventory
replication pipeline	cross-company support and timing	portfolio operations lead	quarterly	transfer assessments and receiving-company plans
external claims and exit evidence	disclosure and diligence readiness	legal, finance and deal teams	before publication or transaction	claims register and evidence room

Frequencies and owners are illustrative management assumptions.

19. Recognise failure modes early

The first failure mode is solution-first investment. A team selects a model or vendor and searches for a problem. The remedy is an approved problem statement, alternatives analysis and accountable benefit owner before technical design.

The second is demo-to-scale compression. A persuasive output becomes evidence of enterprise value. The remedy is a gate between technical feasibility, operating proof, attributable outcome and scale readiness.

The third is metric substitution. Usage, speed or model accuracy replaces the business outcome. The remedy is a theory of change and a value specification that links intermediate and financial measures.

The fourth is missing counterfactual. A favourable result is attributed to AI despite seasonality, training, process change or market movement. The remedy is the strongest feasible evaluation design and an explicit evidence classification.

The fifth is partial cost. Licence cost is compared with a broad labour pool while integration, review, governance and change are excluded. The remedy is a full-cost standard approved by finance.

The sixth is adoption theatre. Users are encouraged to log in while the workflow, incentives and decision rights remain unchanged. The remedy is task-level observation, outcome measurement and authority to redesign work.

The seventh is portfolio double counting. Several use cases claim the same revenue or cost result. The remedy is an overlap register and finance reconciliation.

The eighth is governance by template. Companies complete forms without changing decisions or controls. The remedy is named accountability, evidence-linked gates and board challenge.

The ninth is silent model change. A vendor update alters performance or data handling after approval. The remedy is change notification, monitoring and reassessment triggers.

The tenth is external overclaim. Marketing or transaction materials outrun the proof. The remedy is a claims register tied to the evidence room and disclosure review.

20. Build a reusable sponsor capability

The sponsor's durable advantage lies in a repeatable way to identify, test, govern and transfer AI value. The capability combines investment judgement, operating expertise, evaluation, finance, data, technology, security, risk, legal analysis and change leadership. It should remain proportionate to portfolio size and strategy.

A central methods team can maintain the taxonomy, evidence standard, templates, experiment support, shared vendor diligence, control library and pattern catalogue. Sector specialists can adapt the approach to manufacturing, financial services, healthcare, consumer, business services, infrastructure and technology. Portfolio-company leaders bring the process and customer knowledge required to make the intervention real.

The pattern catalogue should record prerequisites, value mechanism, design, control profile, transfer cost, evidence and failure modes. It should include rejected and retired patterns. A catalogue that contains only successful cases creates overconfidence.

Capability metrics should measure decision quality and realised value. Useful measures include time from intake to decision, share of pilots with a pre-defined baseline, proof packets passing independent challenge, cost of failed tests before scale, replication success, finance-verified benefit, material incidents and retirement speed. Training completion alone is insufficient.

The sponsor should also maintain external scanning. Standards, regulation, vendor capability and threat patterns are changing. NIST was revising AI RMF 1.0 and released a concept note for a critical-infrastructure profile in April 2026 [1]. The European Commission's implementation timeline changed through the 2026 AI Omnibus [4]. Governance should be designed for controlled updates.

21. Approve scale through six independent questions

The scale decision should answer six questions. First, is the use case solving a material problem through a clear value mechanism? Second, does the evaluation support an attributable outcome against a credible counterfactual? Third, is adoption sufficient and is the operating process sustainable? Fourth, does the full-cost bridge produce an acceptable return and cash profile? Fifth, are legal, regulatory, data, model, vendor, cyber, workforce and customer risks controlled? Sixth, can the company and sponsor sustain monitoring, support, assurance and change at the proposed scale?

A weakness in one dimension should remain visible. Strong economics do not remove a legal or safety issue. Strong technical performance does not prove cash value. Complete governance documentation does not create user adoption. The committee can approve conditional scale with defined exposure, evidence and review dates.

The approval record should identify the decision, scope, conditions, accountable owners, funding, expected benefit, evidence class, risk acceptance, monitoring, stop conditions and next review. It should state which facts are company-specific and which patterns came from elsewhere in the portfolio.

Proof before scale is a way to deploy faster with a reliable decision architecture. The portfolio gains a common language for value and risk. Companies retain accountability for their context. Strong use cases receive capital and replication support. Weak assumptions are exposed while the cost of change remains manageable. Boards receive an evidence trail that connects AI activity to operating and financial outcomes.

The result is a portfolio capability that can compound. Each test improves the pattern library, measurement infrastructure, vendor knowledge and control design. The sponsor can direct resources toward interventions with the strongest combination of material value, credible evidence, controlled risk and transferability.

References

- [1] National Institute of Standards and Technology, AI Risk Management Framework <https://www.nist.gov/itl/ai-risk-management-framework>
- [2] National Institute of Standards and Technology, AI RMF Playbook and Resources <https://www.nist.gov/itl/ai-risk-management-framework/ai-risk-management-framework-resources>
- [3] European Union, Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence <https://eur-lex.europa.eu/eli/reg/2024/1689/oj?locale=en>
- [4] European Commission, AI Act regulatory framework and implementation timeline <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>
- [5] National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf>
- [6] European Data Protection Board, Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models https://www.edpb.europa.eu/documents/opinion-of-the-board-art-64/opinion-282024-on-certain-data-protection-aspects-related-to_en
- [7] OECD, OECD AI Principles <https://www.oecd.org/en/topics/ai-principles.html>
- [8] U.S. Government Accountability Office, Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities <https://www.gao.gov/products/gao-21-519sp>
- [9] U.S. Securities and Exchange Commission, SEC Charges Two Investment Advisers with Making False and Misleading Statements About Their Use of Artificial Intelligence <https://www.sec.gov/newsroom/press-releases/2024-36>
- [10] Financial Conduct Authority, Our approach to AI <https://www.fca.org.uk/firms/innovation/ai-approach>
- [11] UK Government, Guidance for using the AI Management Essentials tool <https://www.gov.uk/government/consultations/ai-management-essentials-tool/guidance-for-using-the-ai-management-essentials-tool>
- [12] HM Treasury, The Green Book 2026 <https://www.gov.uk/government/publications/the-green-book-appraisal-and-evaluation-in-central-government/the-green-book-2026>

- [13] HM Treasury, The Magenta Book <https://www.gov.uk/government/publications/the-magenta-book>
- [14] UK Government, Guidance on the impact evaluation of AI interventions <https://www.gov.uk/government/publications/the-magenta-book/guidance-on-the-impact-evaluation-of-ai-interventions-html>
- [15] UK Government, Test and Learn guidance <https://www.gov.uk/government/publications/the-magenta-book/test-and-learn-html>
- [16] UAE Office for Artificial Intelligence, AI Ethics Principles and Guidelines <https://ai.gov.ae/wp-content/uploads/2023/03/MOCAI-AI-Ethics-EN-1.pdf>
- [17] Central Bank of the UAE, Guidance Note on Consumer Protection and Responsible Adoption and Use of Artificial Intelligence <https://rulebook.centralbank.ae/en/rulebook/guidance-note-consumer-protection-and-responsible-adoption-and-use-artificial-intelligence>
- [18] Central Bank of the UAE, SCA, DFSA and FSRA, Guidelines for Financial Institutions Adopting Enabling Technologies <https://rulebook.centralbank.ae/en/rulebook/guidelines-financial-institutions-adopting-enabling-technologies>
- [19] International Organization for Standardization, ISO/IEC 42001 Artificial intelligence management system <https://www.iso.org/standard/42001>
- [20] International Organization for Standardization, Artificial intelligence standards overview <https://www.iso.org/sectors/it-technologies/ai>
- [21] The Institute of Internal Auditors, The IIA's Updated AI Auditing Framework <https://www.theiia.org/en/content/tools/professional/2023/the-iias-updated-ai-auditing-framework>
- [22] Infocomm Media Development Authority, Model AI Governance Framework for Generative AI <https://www.imda.gov.sg/resources/press-releases-factsheets-and-speeches/factsheets/2024/gen-ai-and-digital-foss-ai-governance-playbook>
- [23] Infocomm Media Development Authority, Artificial Intelligence programmes and governance resources <https://www.imda.gov.sg/AI>
- [24] Government of Canada, Algorithmic Impact Assessment tool <https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html>
- [25] Government of Canada, Directive on Automated Decision-Making <https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592>
- [26] Australian Government Department of Industry, Science and Resources, Guidance for AI Adoption Practices <https://www.industry.gov.au/publications/voluntary-ai-safety-standard>
- [27] Australian Securities and Investments Commission, REP 798 Beware the gap: Governance arrangements in the face of AI innovation <https://asic.gov.au/regulatory-resources/find-a-document/reports/rep-798-beware-the-gap-governance-arrangements-in-the-face-of-ai-innovation/>
- [28] Dubai Financial Services Authority, AI survey on adoption and governance in the DIFC <https://www.dfsa.ae/news/new-dfsa-ai-survey-generative-ai-adoption-has-nearly-tripled-within-difc-in-last-12-months-governance-continues-develop>
- [29] International Organization of Securities Commissions, Artificial Intelligence in Capital Markets: Use Cases, Risks, and Challenges <https://www.iosco.org/library/pubdocs/pdf/IOSCOPD788.pdf>
- [30] IFRS Foundation, IAS 38 Intangible Assets <https://www.ifrs.org/content/dam/ifrs/publications/pdf-standards/english/2021/issued/part-a/ias-38-intangible-assets.pdf?bypass=on>

About the Author

Chennakeshav Adya is an independent researcher and Managing Partner at Matchpoint Partners. His research focuses on corporate finance, transactions, governance, strategy and operating models that connect capital decisions with measurable execution.