Hafsa Research and Analysis company

FLAGSHIP WHITE PAPER

Governed Intelligence

Deploying AI Controls Across Close, Audit & FP&A

A practical framework for finance leaders who want the speed of AI without losing control of the numbers.

Edition 1.0 · November 2026

For CFOs, Controllers, Chief Audit Executives, FP&A leaders and Audit Committee members

4 layer governance framework	**5** level autonomy ladder	**3** domains: Close, Audit, FP&A	**4** worked examples with real arithmetic

Contents

Table 1. *What each section covers and who should read it first*

Section	What it covers	Read first if you are
Executive briefing	The argument, the framework and the next 30 days on one page	Everyone
1 The governance gap	Why finance needs AI-specific controls now	CFO, Audit Committee
2 The framework	Four layers, six principles, named roles	Everyone
3 Autonomy and risk tiering	Five-level autonomy ladder and three-tier model	CFO, CRO, Controller
4 Financial close	Control points, journal scoring, accrual agent	Controller
5 Audit	Three lines, audit lifecycle, auditing AI, ECL example	Chief Audit Executive
6 FP&A	Governed forecast cycle and a drift example	CFO, Head of FP&A
7 Cross-cutting foundations	Evidence pack, standards mapping, risk register	CRO, CISO, Controller
8 Implementation roadmap	Twelve-month plan, 30-60-90 days, RACI	Programme sponsor
9 Measuring success	KPI set and balanced scorecard	CFO, Audit Committee
10 Audit Committee briefing	Ten questions and five red flags	Audit Committee
Appendices	Register template and control test script	Practitioners

How to use this paper

Read the executive briefing first. Then go to the section for your domain; each domain section ends with a role briefing that lists the decisions that role needs to make. The figures are designed to be lifted into your own policy and training material. The worked examples use illustrative figures so the arithmetic stays visible; they are not benchmarks.

This paper offers general guidance. It is not legal, accounting or audit advice, and the standards it cites apply differently by jurisdiction and entity. Confirm applicability with your legal, compliance and external audit advisers.

Executive briefing

EXECUTIVE BRIEFING

The 60-second version

- **AI is already inside finance.** It sits in reconciliation tools, ERP copilots, audit analytics and forecasting platforms, usually adopted team by team with no common control standard.
- **Finance outputs carry legal weight.** Statutory accounts, internal control assertions, audit opinions and market guidance are not drafts. A control failure here is a misstatement, not a bad first attempt.
- **Governed Intelligence is the answer.** Every AI-assisted step has an accountable owner, a defined autonomy level, retained evidence and independent assurance. Control intensity scales with risk.
- **Governance is what lets you scale.** Auditors, regulators and boards extend trust to AI that can be evidenced. Controls are the licence to operate, not the brake.

The framework in five moves

1. **Inventory and tier.** Find every AI use in finance, including shadow use, and assign Tier 1, 2 or 3 by what the output touches.
2. **Cap the autonomy.** Set a maximum level on the five-level ladder, from manual (L0) to acting within limits (L4). Raise it only on evidence.
3. **Control at the point of use.** Apply the close, audit and FP&A control matrices in this paper, with named owners and evidence by default.
4. **Evidence as a by-product.** Log inputs, outputs, versions, approvals and overrides automatically so an auditor can re-perform the work.
5. **Assure independently and continuously.** Validate before go-live, monitor drift afterwards, and have internal audit test design and operation.

What to do in the next 30 days

- Appoint an AI governance council chaired by the CFO, with second-line and internal audit seats.
- Run an inventory sprint covering licences, vendor features switched on by default, and analyst-built tools.
- Pause new Tier 1 deployments until they are registered and tiered.
- Issue a one-page usage policy: no confidential ledger data into unapproved tools, human sign-off on anything external.
- Pick three pilots, one each in close, audit and FP&A, and baseline their metrics.
- Brief the Audit Committee and your external auditors on the plan before the first pilot goes live.

1 The governance gap

Finance has always adopted technology faster than it has adopted the controls around it. Spreadsheets arrived before spreadsheet-risk policies, and ERP systems before general IT control frameworks. AI is

following the same pattern with one difference: speed. A vendor can ship a new model overnight, a copilot can be enabled for a whole shared-service centre with a licence change, and an analyst can paste a ledger extract into a public chatbot before any policy says they should not.

Three shifts make this wave different from earlier automation:

- **From deterministic to probabilistic.** Traditional software gives the same output for the same input. AI outputs vary with model version, prompt, context and data, and they can degrade silently as conditions change.
- **From advice to action.** Agentic systems do not only suggest; they can prepare and, if permitted, post entries, send requests and trigger workflows. Their permissions become a control matter.
- **From owned to embedded.** Much finance AI arrives as a feature inside platforms the organisation already uses. The control environment can change without a change ticket ever being raised.

1.1 Why traditional controls are not enough

General IT controls and ICFR design assume software that behaves predictably between releases. Table 1 shows where that assumption breaks and what the governance response should be.

Table 2. *Traditional control assumptions versus what AI changes*

Traditional assumption	What AI changes	Governance response
Same input gives the same output	Output varies with model version, prompt, context and sampling	Version-lock models; test against fixed benchmark sets; log inputs and outputs
Changes pass through change management	Vendors update models without notice; prompts and settings alter behaviour	Treat model versions, prompts and thresholds as controlled items; contract for change notification
Logic can be read and walked through	Complex models are hard to explain	Require proportionate explainability; challenge through back-testing and challenger models
Users are authorised individuals	Agents act on behalf of users, often with broad permissions	Give agents their own identities and least-privilege access; enforce SoD technically
Errors are visible in the output	Plausible but wrong output reads as correct	Tie verification to source data; require reviewer sign-off on material outputs
Performance is stable between releases	Accuracy degrades as data and conditions shift (drift)	Define metrics and thresholds; monitor continuously; revalidate on trigger

1.2 Where AI already operates in finance

Table 3. *Typical AI use cases and their risk character*

Domain	Typical use cases	Risk character
Financial close	Auto-matching in reconciliations; journal anomaly scoring; accrual drafting; variance narratives; intercompany matching	Direct effect on the ledger; high ICFR relevance
Audit	Full-population testing; risk scoring; contract and evidence review; working-paper drafting; control-testing automation	Affects the quality of assurance; reliance and scepticism risk
FP&A	Driver-based and machine-learning forecasts; scenario generation; cash forecasting; board-pack commentary	Drives decisions and external guidance; bias, drift and disclosure risk
Shared tools	Copilots; enterprise search; document extraction from invoices and contracts	Data leakage; fabricated content; unmanaged access

1.3 The cost of ungoverned adoption

- **Control deficiencies.** An AI step that influences the ledger without design documentation, testing or evidence is a deficiency waiting to be found.
- **Impaired auditor reliance.** External auditors cannot rely on what cannot be evidenced, so efficiency gains turn into extra audit effort and fees.
- **Data exposure.** Confidential, personal or price-sensitive data can leave the organisation through tools nobody approved.
- **Decision risk.** A biased or drifting forecast steers capital, hiring and guidance for months before anyone notices.
- **Shadow AI.** Where approved tools are slow to arrive, people use their own. The risk does not disappear; it becomes invisible.

1.4 The opportunity

Governance and speed are not opposites. Organisations that can show an auditor the inventory, the validation file and the override log get approval to widen automation. Those that cannot spend their time defending what they have already deployed. Controls are the licence to scale.

KEY TAKEAWAYS

Section 1 in brief

- AI changes the assumptions behind classic finance controls: determinism, change management, explainability and visible errors.
- The risk is highest where AI touches the ledger, disclosures or audit conclusions.
- Ungoverned adoption costs more than governed adoption: in findings, audit effort and decision quality.

2 The Governed Intelligence framework

Governed Intelligence is the practice of deploying AI inside finance processes under explicit, proportionate and evidenced controls, so that every AI-assisted step has an accountable owner, a defined level of autonomy, retained evidence and independent assurance.

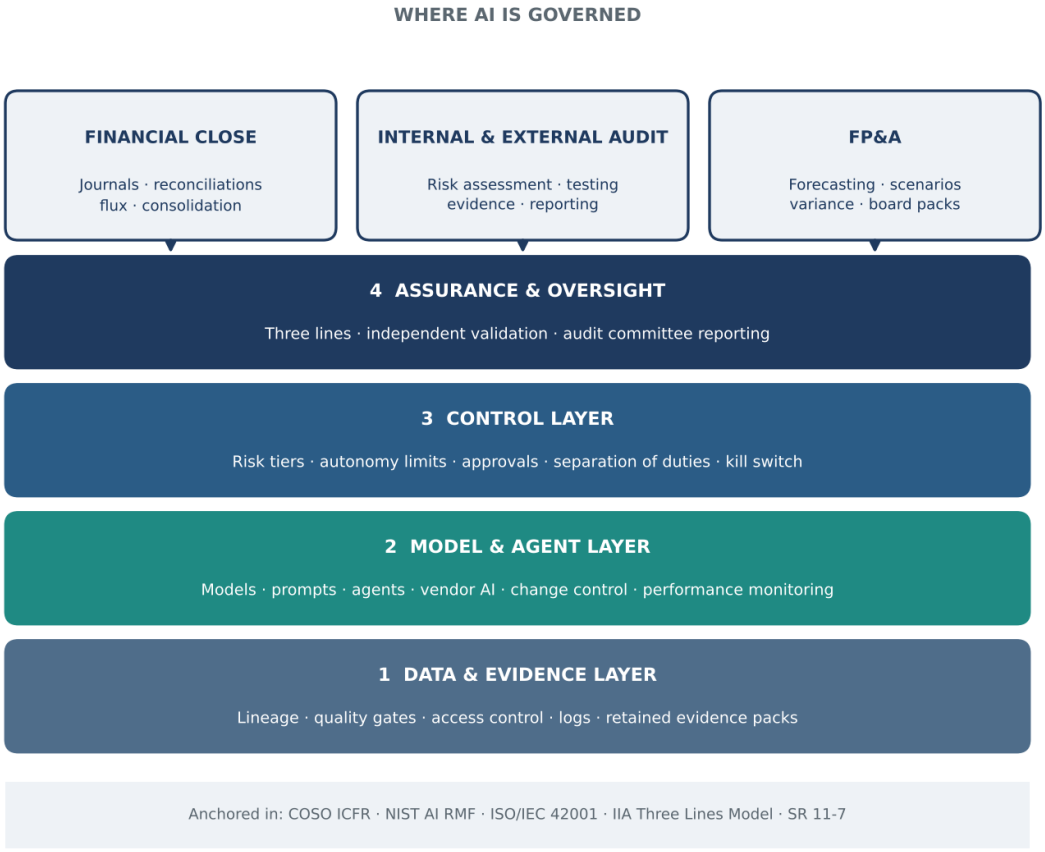


Figure 1. *The four-layer framework applied across close, audit and FP&A.*

2.1 The four layers

Layer 1: Data and evidence

Every AI output is only as reliable as the data behind it, and every control is only as good as the evidence it leaves.

- Data lineage from source system to model input, with quality gates on completeness, validity and timeliness.
- Role-based access, with confidential and personal data classified and ring-fenced.
- Immutable logs of inputs, outputs, versions and approvals, retained to your record-keeping standard.

Layer 2: Model and agent

- A register of every model, prompt, agent and vendor feature, each with an owner and a version.
- Formal change control for models, prompts, parameters and thresholds.
- Performance monitoring against defined metrics: accuracy, bias, drift, override rate.

Layer 3: Control

- Risk tiers and autonomy ceilings set by policy (Section 3).
- Approval gates, separation of duties and technical limits on what agents can do.
- A kill switch and a documented fallback to manual operation.

Layer 4: Assurance and oversight

- Independent validation before go-live for Tier 1 and 2 use cases.

- Second-line monitoring and challenge; internal audit coverage in the annual plan.
- Regular reporting to the Audit Committee on inventory, performance and incidents.

2.2 Six design principles

Table 4. *Principles and the question each one answers*

#	Principle	In practice	Test question
1	Accountable human	A named person owns each use case and each material output	Who signs for this output?
2	Proportionate control	Control intensity scales with risk tier, not with enthusiasm	Why is this the right tier?
3	Explainable enough to challenge	Reviewers can see drivers, sources and confidence	Could a reviewer disagree for a reason?
4	Evidenced by default	Evidence is a by-product of the workflow	Could an auditor re-perform this?
5	Separation of duties preserved	No AI actor both prepares and approves its own work	Who approves what the AI prepared?
6	Continuously monitored	Performance, drift, overrides and incidents are measured	How would we know it had gone wrong?

2.3 Roles that make it work

- **Use-case owner (first line).** Accountable for purpose, outputs, performance and control operation. Cannot delegate this to a vendor.
- **Control owner.** Operates and evidences the specific controls in the matrices.
- **Model validator (second line).** Independent of the builder; tests design, data and performance before go-live and on change.
- **Data steward.** Owns data quality, lineage and access for the inputs.
- **AI governance council.** Chaired by the CFO; approves Tier 1 use cases, sets the autonomy ceilings and reviews monitoring.
- **Internal audit (third line).** Provides independent assurance on design and operating effectiveness.

3 Autonomy and risk tiering

Two decisions set the control intensity for any AI use case: how much the AI is allowed to do (autonomy) and how much damage a failure could cause (risk tier). Keeping them separate prevents the common error of treating a low-risk use at high autonomy, or a high-risk use at low autonomy, as if both needed the same controls.

3.1 The autonomy ladder

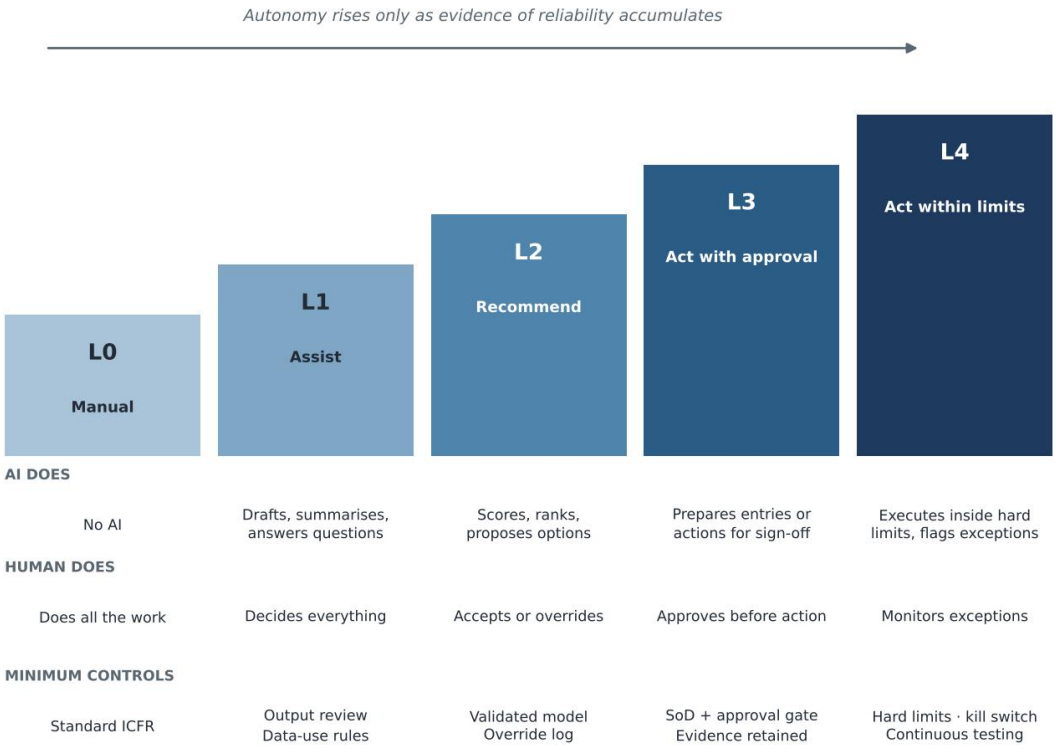


Figure 2. *Five autonomy levels with the AI role, human role and minimum controls at each.*

The ladder describes behaviour, not technology. The same model can run at L1 in one process and L3 in another. Two rules apply:

- **Autonomy is earned.** A use case moves up one level at a time, and only on documented evidence: back-test results, override rates and the sign-off of the owner and the second line.
- **Autonomy is revocable.** Defined triggers, such as an override rate or an error threshold, drop the use case back a level automatically.

3.2 Risk tiering

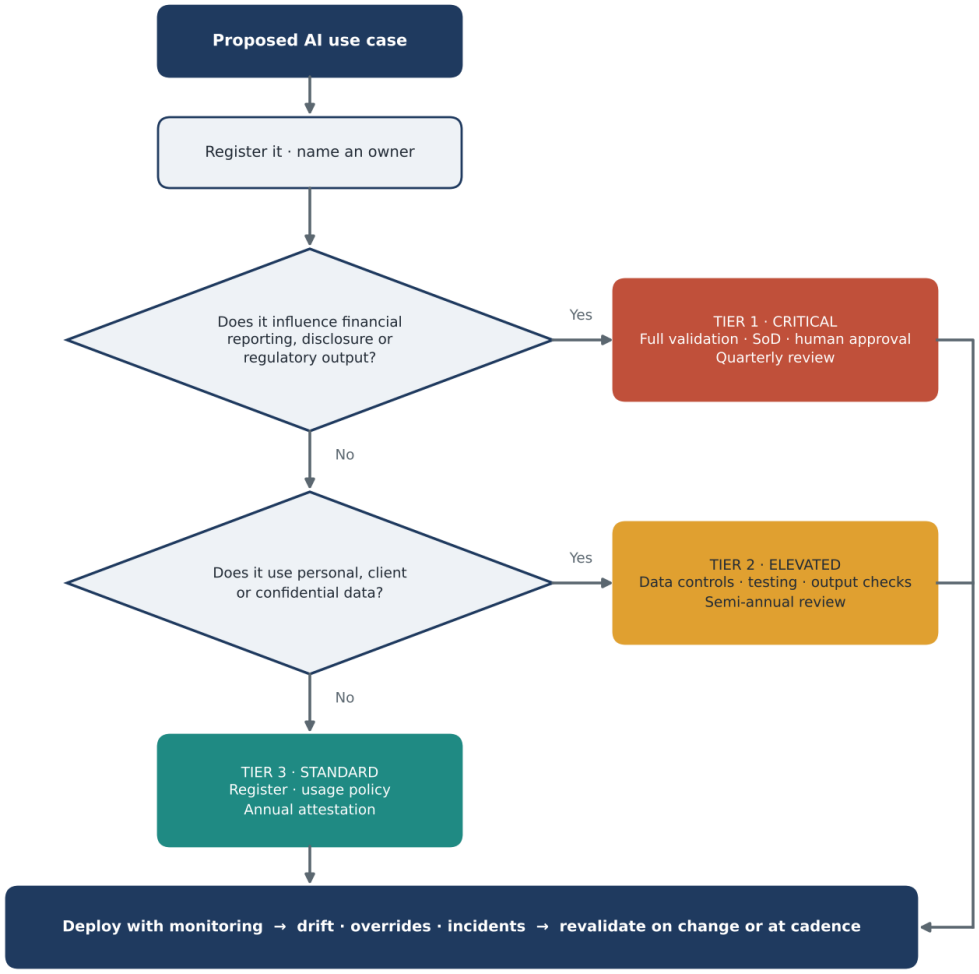


Figure 3. *Decision flow for assigning a risk tier to a proposed AI use case.*

Table 5. *Risk tiers, minimum controls and review cadence*

Tier	Definition	Examples	Minimum controls	Review
1 Critical	Influences financial statements, disclosures, regulatory returns or audit conclusions	ECL inputs; accrual agent; consolidation mapping; audit-evidence analytics	Independent validation before go-live; documented design; SoD enforced; human approval; evidence pack	Quarterly and on change
2 Elevated	Handles personal, client or confidential data, or informs management decisions with no direct ledger effect	Forecasting; contract review; management variance commentary	Data-handling controls; output testing; reviewer check; documented limitations	Semi-annual
3 Standard	Productivity use with no sensitive data and no direct decision input	Meeting summaries; drafting internal notes	Usage policy; register entry; annual attestation	Annual

Autonomy ceilings by tier. Tier 1 use cases are capped at L3 (act with approval); L4 is allowed only for narrowly bounded, reversible actions with hard limits and a documented exception from the council. Tier 2 is capped at L3. Tier 3 may run to L4. These ceilings are defaults; a council can set lower ones for specific processes.

4 Governing AI in the financial close

The close is the most control-dense process in finance, which makes it both the best place to find value and the place where AI errors are most expensive. The approach is to identify each point where AI touches the process and attach a specific, evidenced control to it. Figure 3 maps six typical points.

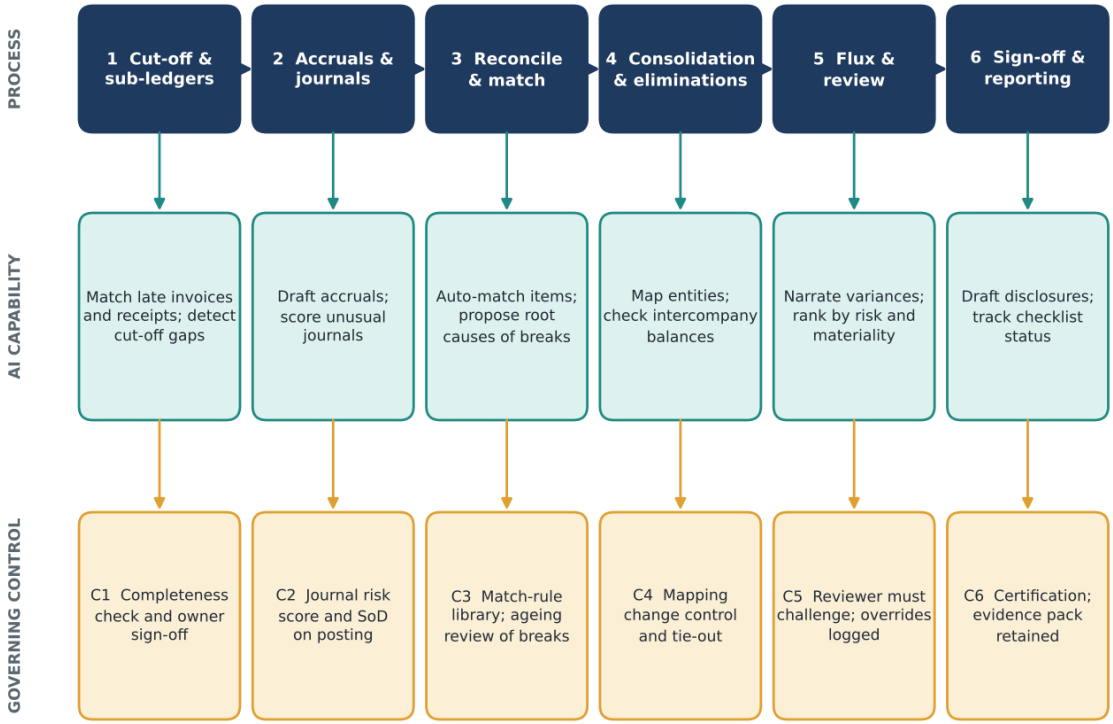


Figure 4. *Close stages, the AI capability at each stage and the governing control.*

4.1 Close control matrix

Table 6. *Controls C1 to C6 for AI in the close*

ID	Risk	AI-enabled step	Control	Evidence	Owner, frequency
C1	Late invoices or receipts missed	Matching engine flags unrecorded items	Completeness check against cut-off report; owner sign-off	Exception list; sign-off	AP lead, each close
C2	Erroneous or inappropriate journals	Journal scoring; draft accruals	Score threshold; every high-score journal reviewed; SoD on posting	Score log; reviewer notes	GL manager, each close
C3	Unreconciled or ageing breaks	Auto-match; root-cause suggestions	Approved match rules; tolerance limits; ageing review	Rule version; break ageing report	Recon owner, monthly
C4	Wrong entity mapping or eliminations	Mapping suggestions; intercompany checks	Mapping change control; tie-out to trial balance	Change log; tie-out	Group reporting, each close
C5	Misleading variance explanation	Generated flux commentary	Reviewer verifies drivers to source; overrides logged	Source links; override log	Controller, each close
C6	Incomplete close certification	Checklist tracking; draft disclosures	Sub-certifications; evidence pack complete before sign-off	Evidence pack index	CFO delegate, each close

4.2 Example 1: Journal anomaly scoring

WORKED EXAMPLE

Is the control working?

A group posts 18,400 journals in a month-end close. A model scores each from 0 to 1 using amount, timing, account combination, user behaviour and narrative features. Policy sets the review threshold at 0.80.

Step	Result	Governance check
Population scored	18,400 journals	Journals scored equal journals in the ledger extract
Flagged at 0.80 or above	552 (3.0%)	Threshold approved by the Controller; changes logged
Confirmed issues after review	46	Precision = 46 ÷ 552 = 8.3%
Seeded-error test	17 of 20 detected	Recall = 85%; minimum 80% to stay in service
Risk-based sample of unflagged	60 of 17,848, no exceptions	Tests for blind spots outside the flagged set

Reading the result. A precision of 8.3% looks low, but the control cuts review from 18,400 journals to 552 (3.0%) while still catching most seeded errors. Recall, not precision, is the number that protects you. If recall falls below 80% in the next quarterly seeded test, treat the control as deficient, return the model to validation and revert to the prior manual sampling until it passes.

4.3 Example 2: An accrual agent at L3

WORKED EXAMPLE

Prepare, do not post

An agent drafts month-end accruals for utilities, freight and professional fees from contracts, purchase orders and goods-received data. It runs at L3: it prepares, a human approves. The guardrails turn that sentence into controls:

- **Rights.** The agent has its own identity with prepare rights only. It has no posting right, so it cannot approve its own work.
- **Limits.** It may draft entries up to USD 50,000 each and USD 400,000 in total per close. Anything above escalates to the GL manager.
- **Traceability.** Each draft carries references to the contract, PO and receipt it relied on, and an automatic reversal date.
- **Monitoring.** If approvers change more than 15% of drafts in a close, the agent drops to L2 until the cause is fixed.
- **Fallback.** A kill switch disables the agent; the close checklist includes a manual accrual procedure.

CONTROLLER BRIEFING

What the Controller should decide this quarter

- Which of the six close controls are live today, and which exist only in the vendor brochure.
- The threshold, seeded-test cadence and recall minimum for journal scoring.
- Whether any agent holds both prepare and post rights anywhere in the close.
- Who owns the evidence pack at sign-off, and where it is stored.

5 Governing AI in audit and assurance

AI touches audit in two directions. It is a tool that auditors use to test more of the population, and it is a subject that auditors must examine when management deploys it. Both need discipline. The first protects the quality of the audit; the second protects the quality of the financial statements.

5.1 The three lines and AI

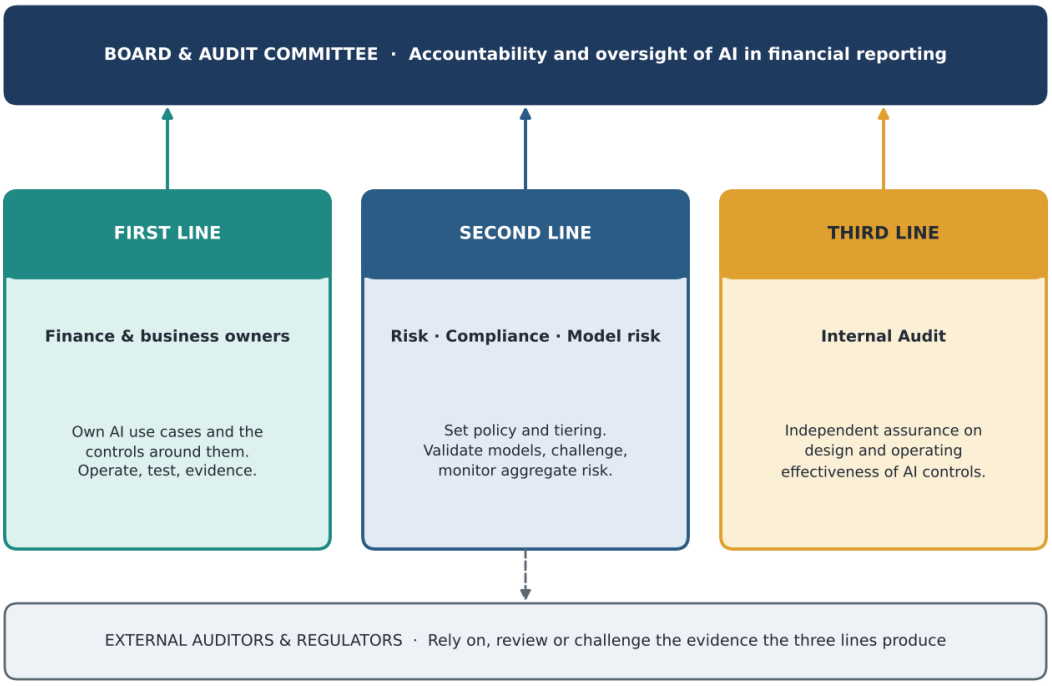


Figure 5. *Responsibilities of the three lines for AI in financial reporting.*

The most common failure is a first line that treats validation as the vendor's job. A vendor can supply performance evidence, but accountability for fitness for purpose in your process, with your data, stays with the use-case owner.

5.2 AI inside the audit lifecycle

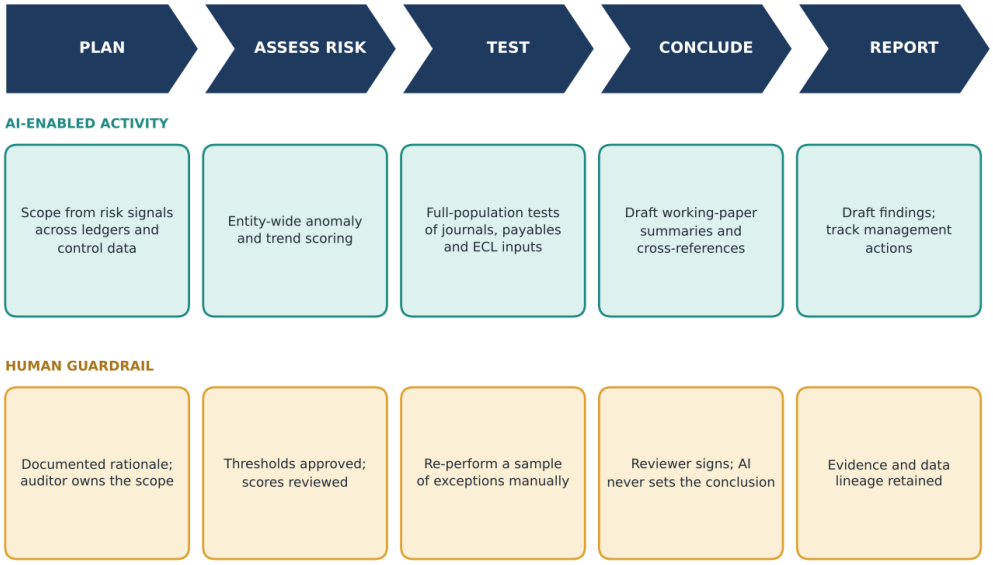


Figure 6. *AI-enabled activities at each audit stage with the guardrail that keeps the auditor in charge.*

Full-population testing changes the shape of audit work: instead of sampling and extrapolating, the auditor tests every item against a rule and then investigates exceptions. That is a real quality gain, but it creates three risks that need explicit controls:

- **False assurance.** A clean run proves only that the tests found nothing; it does not prove the tests were the right ones. Seed errors and re-perform a sample of exceptions manually.
- **Automation bias.** Reviewers tend to accept confident machine output. Require reviewers to record why they accepted, not only that they did.
- **Evidence and confidentiality.** Client data processed by AI tools must stay inside approved environments, and the evidence trail (data snapshot, script, version, output) must be retained.

5.3 Auditing the AI itself

Table 7. *An audit programme for AI controls in financial reporting*

Area	Audit question	Evidence to request
Inventory and tiering	Is every AI use in scope identified and tiered correctly?	Register; tiering rationale; reconciliation to licences and vendor feature lists
Governance	Who approved each Tier 1 use case, and on what basis?	Council minutes; approval records; owner attestations
Data	Are inputs complete, accurate and appropriately restricted?	Lineage record; quality-gate results; access listings
Validation	Was the model independently tested before use and after material change?	Validation report; back-test; challenger comparison
Change control	Are model, prompt and threshold changes approved before use?	Configuration log; change tickets; vendor release notices
Access and SoD	Can any single identity, human or agent, prepare and approve?	Role matrix; service-account listing; permission test

Area	Audit question	Evidence to request
Monitoring	Are accuracy, drift and overrides tracked, with action on breaches?	Monitoring reports; breach log; remediation records
Third parties	Do contracts cover data use, change notice, audit rights and exit?	Contracts; vendor assurance reports; exit plan

5.4 Working with external auditors

External auditors are expected to understand how an entity uses IT, including automated tools, in its system of internal control (see ISA 315, Revised 2019). Early transparency reduces friction and cost:

- Share the inventory, tiering and validation summaries before year-end planning, not during fieldwork.
- Identify which AI outputs feed the financial statements, and which controls the auditor might wish to rely on.
- Make evidence retrievable without a bespoke data-extraction project.
- Agree in advance how changes to models during the year will be communicated.

5.5 Example 3: An AI update to an IFRS 9 ECL input

WORKED EXAMPLE

When the model says the loss rate has moved

A portfolio of trade receivables has Stage 1 exposure at default (EAD) of 5,000,000. The current 12-month probability of default (PD) is 1.8% and loss given default (LGD) is 45%. A machine-learning credit model using behavioural and macroeconomic signals proposes a PD of 2.4%.

Item	Current	AI-proposed	Change
PD	1.80%	2.40%	+0.60 points
LGD	45%	45%	None
EAD	5,000,000	5,000,000	None
ECL = PD × LGD × EAD	40,500	54,000	+13,500 (+33.3%)

Governance walk-through.

1. **Trigger.** A movement beyond the approved ±10% threshold routes the change to the Tier 1 review committee; it cannot flow to the ledger automatically.
2. **Explainability.** The model owner presents the main drivers (for example, the unemployment outlook and delinquency roll rates), their direction and their size.
3. **Back-test.** Predicted default rates are compared with observed rates for the last eight quarters, against a simpler challenger model.
4. **Judgement.** The committee accepts, rejects or applies a management overlay, and minutes the reasoning.
5. **Disclosure.** The approved movement flows into the ECL reconciliation and sensitivity disclosures.
6. **Evidence.** Model version, data snapshot, validation file and approvals are retained.

Discounting and multiple forward-looking scenarios are omitted to keep the arithmetic visible. The control logic is unchanged when they are included.

CAE BRIEFING

What the Chief Audit Executive should decide this year

- Whether AI governance in financial reporting is on the audit plan, and with what skills.
- How internal audit will test its own use of AI, and keep evidence of review.
- Which AI controls the external auditor may rely on, and whether internal audit can support that.
- Where independence is at risk: for example, internal audit advising on the design of controls it will later test.

6 Governing AI in FP&A

Forecasts steer capital, hiring and guidance to the market. AI can sharpen them by finding patterns in drivers that analysts cannot hold in their heads, but a forecast that is wrong in a consistent direction does more harm than one that is merely noisy. FP&A governance therefore concentrates on bias, drift, explainability and the discipline of human challenge.

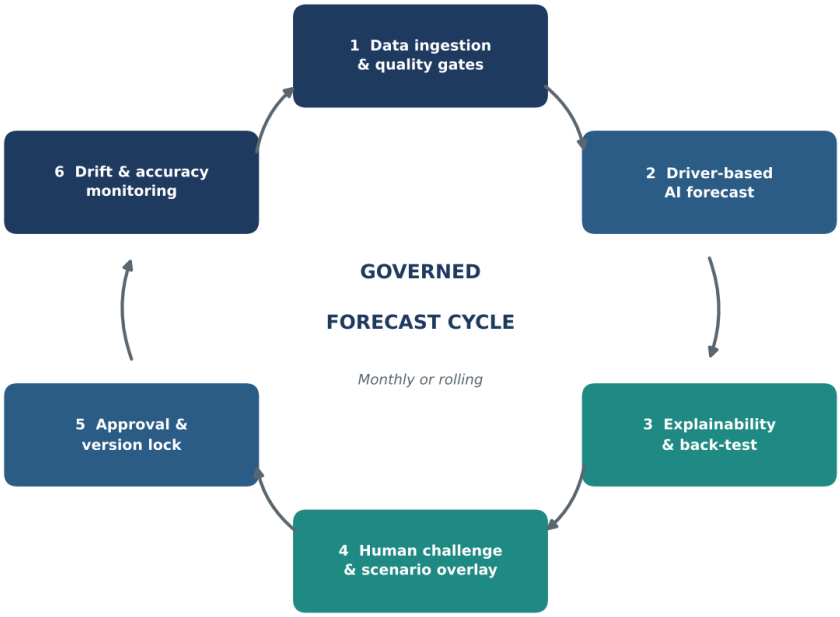


Figure 7. *The governed forecast cycle, from data quality gates to drift monitoring.*

6.1 Controls across the cycle

Table 8. *Controls for each stage of the governed forecast cycle*

Stage	Control	What it prevents
1 Data ingestion	Quality gates on completeness, timeliness and reconciliation to the ledger	Forecasting from incomplete or stale actuals

Stage	Control	What it prevents
2 AI forecast	Approved model version and parameters; documented drivers and assumptions	Silent changes in method; undocumented overrides
3 Explainability and back-test	Driver attribution; comparison with prior-period accuracy and a naive baseline	Black-box numbers nobody can challenge
4 Human challenge	Named reviewer challenges the top drivers and applies scenario overlays; overlays logged	Automation bias; unlogged judgement
5 Approval and version lock	Approved forecast locked and versioned; changes need re-approval	Version sprawl; numbers that change after sign-off
6 Drift monitoring	Accuracy and bias rules with action thresholds	Slow degradation going unnoticed

6.2 Example 4: Catching forecast drift

WORKED EXAMPLE

Two rules, two different warning dates

A company uses an AI model for one-month-ahead revenue forecasts. FP&A tracks absolute percentage error (APE = |actual − forecast| ÷ actual) each month and applies two rules approved by the council: **Rule A (accuracy):** APE above 10% for two consecutive months. **Rule B (bias):** four consecutive months with errors in the same direction.

Month	Actual (USD m)	Forecast (USD m)	Error direction	APE
June	40.5	40.0	Under	1.2%
July	42.0	41.2	Under	1.9%
August	44.5	43.1	Under	3.1%
September	47.0	44.3	Under	5.7%
October	46.0	41.2	Under	10.4%
November	48.0	42.6	Under	11.3%

What the rules show. Mean absolute percentage error across the six months is 5.6%, which looks acceptable on its own. Rule B trips in September, after four consecutive under-forecasts. Rule A does not trip until November, when APE has exceeded 10% two months running. Over the period the model under-forecast actual revenue by USD 15.6m in total, or 5.8% of the USD 268.0m actually earned.

Response. Stop publishing the model's forecast as the base case, run the forecast at L1 (analyst-built, with the model as a challenger), investigate the cause (for example, a shift in customer mix the training data did not contain), retrain and revalidate before restoring autonomy. The bias rule bought two months of earlier warning, which is why accuracy metrics alone are not enough.

6.3 Forward-looking information and disclosure

AI-generated commentary and forecasts can flow into earnings guidance, board packs and lender reporting. Route anything external through the existing disclosure controls: the disclosure committee reviews it, the numbers tie to approved sources, and no forward-looking statement is released on the

strength of a model output alone. Keep a record of what was generated, what was changed and who approved it.

CFO BRIEFING

What the CFO should decide this quarter

- The accuracy and bias rules for each forecast model, and who acts when they trip.
- Which forecasts may be published at L2 or L3, and which stay human-built with AI as a challenger.
- Who owns the approved version of the forecast, and how changes after lock are controlled.
- How AI-assisted commentary is reviewed before it leaves the building.

7 Cross-cutting foundations

7.1 The minimum evidence pack

An evidence pack lets an independent person understand, re-perform and challenge an AI-assisted result. Generate it from the workflow rather than assembling it afterwards.

Table 9. *Minimum evidence pack for a Tier 1 use case*

Artefact	Content	Owner
Use-case record	Purpose, owner, tier, autonomy level, data sources, known limitations	Use-case owner
Design and validation file	Method, assumptions, test results, challenger comparison, approval	Model validator
Data lineage record	Sources, transformations, quality checks, snapshot reference	Data steward
Configuration log	Model version, prompts, parameters and thresholds with change history	Technology owner
Run log	Inputs, outputs, timestamps, user or agent identity	Technology owner
Review and override record	Reviewer, decision and reason	Control owner
Monitoring report	Accuracy, drift, override rate, incidents	Second line
Incident and exception log	Event, impact, root cause, remediation	Use-case owner

7.2 Mapping to recognised frameworks

Table 10. *How this paper aligns with established frameworks*

Framework	What it contributes	Where it appears
COSO Internal Control: Integrated Framework (2013)	Control environment, risk assessment, control activities, information and communication, monitoring	Sections 3 to 6; control matrices
SOX 404 and PCAOB AS 2201 (where applicable)	Management's assessment of ICFR and auditor attestation	Close controls; evidence pack
NIST AI Risk Management Framework 1.0	Govern, Map, Measure and Manage functions for AI risk	Layers 1 to 4; monitoring
ISO/IEC 42001:2023	Management-system standard for AI: policies, roles, continual improvement	Council; documented procedures
IIA Three Lines Model (2020)	Roles for management, risk functions and internal audit	Figure 5

Framework	What it contributes	Where it appears
SR 11-7 Model Risk Management (US supervisors, 2011)	Model development, validation, ongoing monitoring and governance	Tier 1 validation; Example 3
IFRS 9 Financial Instruments	Expected credit loss requirements	Example 3
EU AI Act	Risk-based regulation of AI systems	Tiering logic; check applicability by use

Applicability depends on jurisdiction, sector and entity. Treat this table as a starting map and confirm obligations with legal and compliance.

7.3 Top risks and mitigations

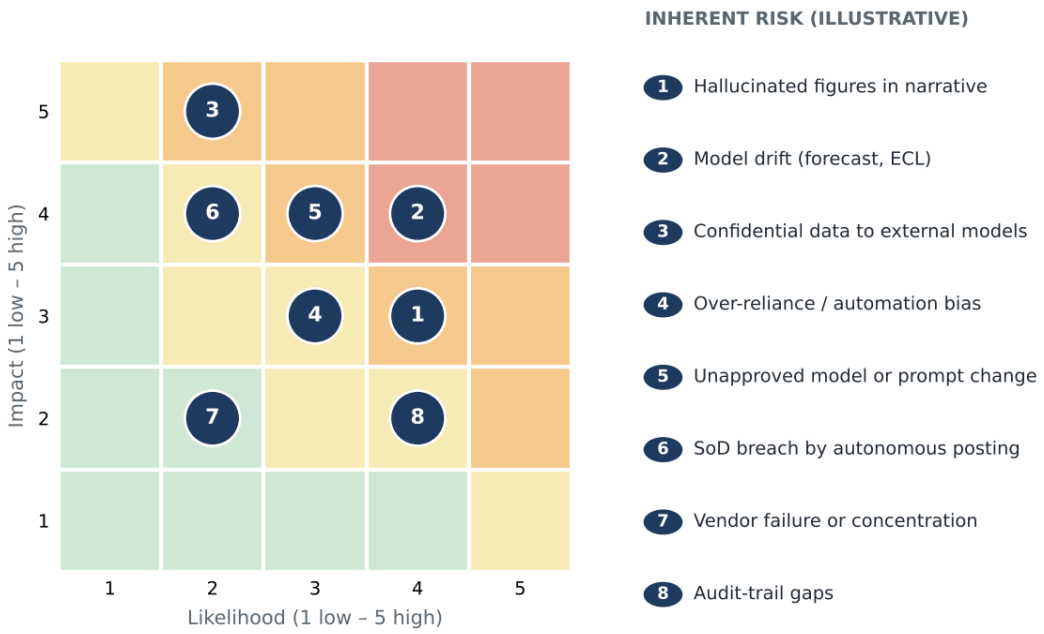


Figure 8. *Inherent risk map for AI in finance (illustrative; calibrate to your own assessment).*

Table 11. *Risk register: the eight risks in Figure 8*

#	Risk	Principal mitigations
1	Hallucinated figures in narrative	Source-linked commentary; reviewer verification; no unreviewed external use
2	Model drift in forecasts and ECL	Accuracy and bias rules; seeded and back-tests; revalidation triggers
3	Confidential data to external models	Approved environments; data classification; contractual data-use limits
4	Over-reliance and automation bias	Recorded acceptance rationale; periodic manual re-performance
5	Unapproved model or prompt change	Configuration control; vendor change notice; release testing
6	SoD breach by autonomous posting	Separate agent identity; prepare-only rights; permission testing
7	Vendor failure or concentration	Exit plan; manual fallback; multi-vendor option for critical uses
8	Audit-trail gaps	Immutable logs; evidence pack completeness check at sign-off

Third-party AI. Embedded vendor AI is the hardest to see. Contracts should secure notice of material model changes, limits on the vendor's use of your data (including for training), audit and assurance rights, incident notification, and an exit plan that preserves your logs.

8 Implementation roadmap

A twelve-month programme moves from foundations, through proof in three pilots, to scaled and independently assured operation. The sequence matters: policy and inventory first, pilots second, scale last.

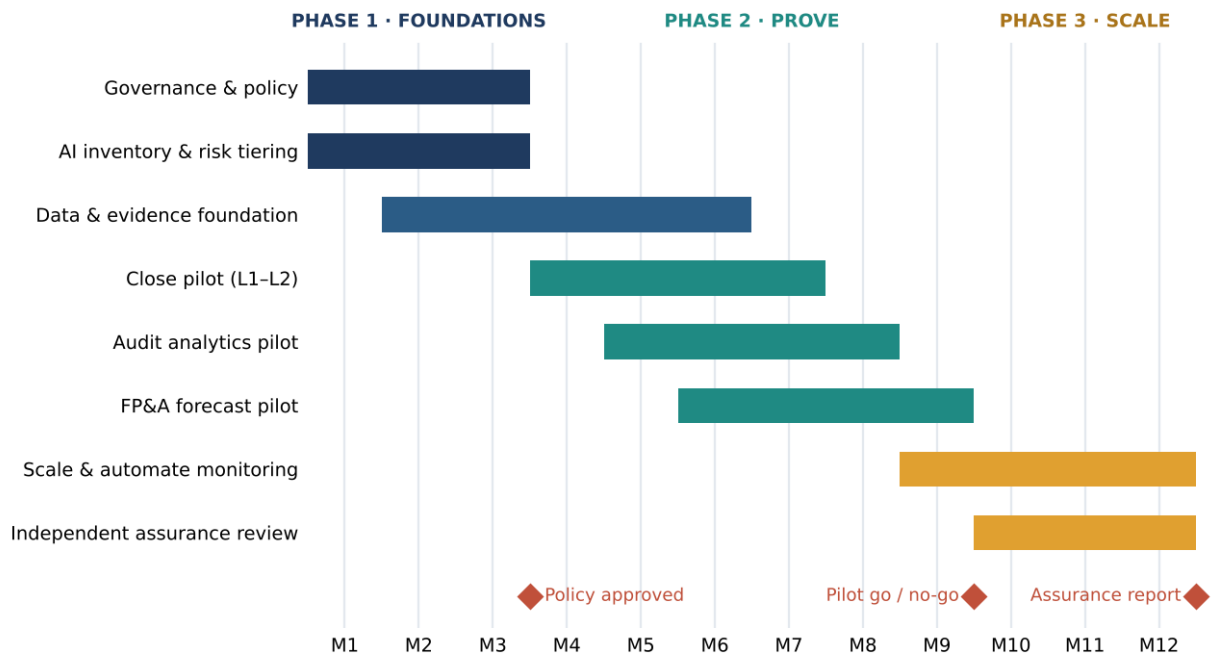


Figure 9. *Twelve-month roadmap with three phases and three milestones.*

8.1 The first 90 days

Table 12. *30-60-90 day plan*

Window	Actions	Outputs
Days 1 to 30	Pause new Tier 1 deployments pending registration; build the inventory, including shadow AI; appoint the council and use-case owners; draft the policy	Inventory v1; council charter; draft policy
Days 31 to 60	Tier the inventory; define the minimum evidence pack; choose three pilots (close, audit, FP&A); baseline metrics	Tiered register; pilot charters with baselines
Days 61 to 90	Approve the policy; implement logging and approval gates for pilots; run the first validation; brief the Audit Committee and external auditors	Approved policy; validated pilots; committee briefing

8.2 Who does what

Table 13. *RACI for AI governance in finance*

Activity	Finance owner	Second line	Internal audit	Technology	Audit Committee
Policy and tiering	C	A / R	C	C	I
Use-case approval	A / R	C	I	C	None
Model validation	C	A / R	I	C	None
Operating controls	A / R	C	I	R	None

Activity	Finance owner	Second line	Internal audit	Technology	Audit Committee
Monitoring and reporting	R	A	I	R	I
Independent assurance	C	I	A / R	I	I
Incident response	A / R	C	I	R	I

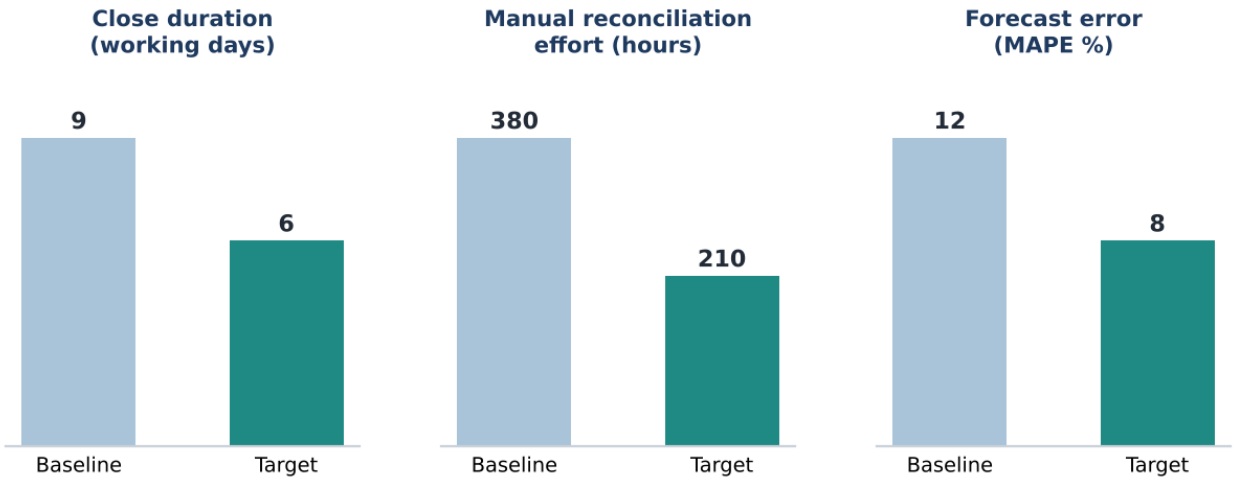
R = responsible, A = accountable, C = consulted, I = informed.

8.3 Operating model

- **Council cadence.** Monthly during the first year, then quarterly. Standing agenda: new use cases, autonomy changes, monitoring exceptions, incidents.
- **Decision rights.** The council approves Tier 1 use cases and any autonomy increase; second line approves Tier 2; owners attest Tier 3.
- **Exception path.** A time-limited exception requires a named risk acceptor and an expiry date.

9 Measuring success

What gets reported gets managed. A balanced scorecard pairs value measures, such as faster closes, with control-health measures, so that speed gains are only counted when controls are green.



Illustrative target model for a mid-size multinational finance function. Not benchmark data.

Figure 10. *Illustrative target outcomes from a governed pilot programme.*

Table 14. *Suggested KPI set*

Category	KPI	Type	Signal to watch
Control effectiveness	Tier 1 use cases with current validation	Leading	100%
	Seeded-error detection rate	Leading	At or above the control minimum (e.g. 80%)
	Override rate on AI recommendations	Leading	Stable band; sudden change investigated
Performance	Forecast MAPE and bias	Lagging	Within approved tolerance
	Auto-match and false-match rate	Lagging	Per reconciliation policy
Evidence and audit	Evidence packs complete at sign-off	Leading	100%

Category	KPI	Type	Signal to watch
	Open AI-related audit findings and age	Lagging	Trending down
Value	Close duration; manual effort hours	Lagging	Targets in the business case
Risk	AI incidents and near misses	Lagging	Reported and root-caused within SLA

Design the dashboard so that the control KPIs sit beside the value KPIs. If a value measure improves while a control measure turns amber, the council discusses the amber one first.

10 Audit Committee briefing

The Audit Committee does not need to understand model internals. It needs to ask questions whose answers reveal whether management knows where AI is used, controls it and can prove it.

AUDIT COMMITTEE BRIEFING

Ten questions to ask management

1. Where is AI used in financial reporting, and who owns each use?
2. How are use cases risk-tiered, and which are Tier 1?
3. What is the maximum autonomy allowed for each, and who approved it?
4. How do we know the models are accurate now, not only at go-live?
5. What happens when a model drifts or a vendor changes it?
6. Can any AI act on the ledger without human approval? Where?
7. What data leaves the organisation, and under what contract?
8. What evidence would an external auditor see, and have they seen it?
9. How have incidents and near misses been handled?
10. What independent assurance has covered this, and what did it find?

Five red flags

- "We use AI, but we have no inventory."
- "The vendor says it is validated."
- "Nobody can say what the override rate is."
- The same service account prepares and posts.
- Evidence is assembled at quarter-end.

Conclusion: earn the right to scale

AI will keep moving into the close, the audit and the forecast. The organisations that benefit most will not be those that adopt first, but those that can show, at any moment, what the AI is doing, who is accountable, how well it is performing and what happens when it fails. Governed Intelligence is that discipline. It turns AI from a risk the board tolerates into a capability the board, the auditors and the regulators can trust.

Appendix A Use-case register template

Table 15. *Fields for the AI use-case register, with an example entry*

Field	Description	Example
ID	Unique reference	AI-CL-014
Name	Short descriptive name	Accrual drafting agent
Process	Close, audit, FP&A or shared	Close
Owner	Accountable individual	Financial Controller
Risk tier	1, 2 or 3 with rationale	Tier 1: influences ledger entries
Autonomy level	L0 to L4 and approved ceiling	L3 (ceiling L3)
Inputs and data	Sources and sensitivity	Contracts, POs, goods received; confidential
Outputs	What it produces and where it goes	Draft accrual journals for approval
Key controls	Control IDs from the matrix	C2, C5, C6
Model and version	Vendor, model, configuration version	Vendor X agent v2.3, config 14
Validation	Date, validator and result	Validated; next review next quarter
Known limitations	Where it should not be used	Not for one-off or non-contract costs

Appendix B Sample control test script

Test script for the journal-scoring control (C2), suitable for internal audit or second-line testing:

1. Obtain the approved threshold, the model version in use and the completeness reconciliation for the period.
2. Re-perform completeness: compare the count of journals scored with the ledger extract and investigate differences.
3. Select 25 flagged journals and confirm each has a reviewer decision evidenced before period-end sign-off.
4. Run a seeded-error test of 20 journals; confirm the detection rate meets the minimum (80%).
5. Select a risk-based sample of unflagged journals and test for errors the model missed.
6. Inspect the change log: every model, threshold or prompt change approved before it was used.
7. Confirm access: no single human or service identity can both prepare and approve a journal.
8. Conclude on design and operating effectiveness; document exceptions, root causes and remediation owners.