Model Risk at Credit Committee: When AI Scores May Inform a Lending Decision

Evidence Thresholds for Data Quality, Explainability, Stability, Overrides and Realised Performance

Authored by Chennakeshav Adya, Independent Researcher

September 2026

Abstract

AI and other quantitative scores can rank borrowers, estimate default risk, forecast loss and prioritise credit review. Their apparent precision can exceed the strength of the underlying data, assumptions and validation. A committee that receives a score without purpose, uncertainty, limitations and decision boundaries may over-rely on the model or reject useful evidence without a disciplined basis.

This paper develops a board-grade framework for deciding when an AI score may inform a lending decision. It connects model inventory and tiering with data provenance, conceptual soundness, testing, independent validation, explainability, stability, calibration, overrides, drift, outcome analysis and human authority. The framework separates model output from contractual facts, credit judgement, legal interpretation and approval. It also covers vendor models, alternative data, generative-AI interfaces, implementation risk and manual fallback.

Five original figures and five decision tables present the model-evidence ladder, decision-rights map, override analysis, drift dashboard and approval-memo architecture. A worked example uses a hypothetical borrower and illustrative assumptions. All amounts, scores, thresholds, probabilities and scenarios are analytical assumptions. The paper does not provide accounting, audit, legal, regulatory, tax, investment, lending, credit, model-validation, technology or risk-management advice.

JEL Classification: G21, G32, C52, C53, M15

Keywords: model risk, credit committee, AI credit scoring, model validation, explainability, overrides, drift, lending decisions, credit governance

1. Begin with the credit decision

A model exists to support a defined decision. The committee may be considering approval, structure, limit, pricing, covenant, collateral, monitoring, classification or another authorised action. Each use has a different consequence and evidence threshold.

The decision record should identify the facility, borrower, exposure, requested authority, model purpose, score date, population, data cut-off and accountable owners. It should distinguish model output from observed facts, contractual terms, forecasts, judgement and legal conclusions.

A score can inform review without determining the decision. The committee remains responsible for evaluating repayment capacity, structure, downside, evidence quality, limitations and alternatives within its delegated authority.

Decision consequence should drive the required assurance. A score used to order a work queue can tolerate different uncertainty from one that affects approval, pricing, limit, covenant or collateral. The paper should show the maximum influence permitted, the evidence supporting that

influence and the decision elements reserved for accountable judgement.

Timing also matters. A model result may be current at calculation and stale by committee after new accounts, liquidity movements, customer events or market changes. The pack should state the effective date, intervening evidence and whether re-scoring or independent analysis is required before reliance.

Material conflicts need a defined route. If the model indicates low risk while cash-flow analysis shows stress, the committee should investigate data, assumptions, horizon and borrower context. It should not average incompatible conclusions into a false compromise. The minutes should record how the conflict was resolved and which evidence governed the action.

2. Define the model and the model system

A model transforms inputs through assumptions and methods into an estimate used in decision-making. The surrounding system can include source feeds, transformations, rules, third-party services, user adjustments, interfaces and downstream calculations.

Risk can arise from the mathematical method, data, implementation, use or interaction among components. Inventory and governance should cover the complete decision chain rather than only the central algorithm.

Deterministic tools can also have material impact. Complex rules, scorecards and calculation engines may require model-like controls according to policy and consequence.

The system boundary should include manual inputs and downstream use. A sound score can become unsafe when a user selects an unsupported value, an interface maps fields incorrectly or another system converts a rank into a limit. Validation should therefore test the production decision path rather than an isolated model file.

Dependencies should be inventoried. Several credit tools may share borrower data, macroeconomic scenarios, identity services, vendors or feature pipelines. A common failure can affect many decisions simultaneously. Aggregate reporting should show concentrations and compensating controls.

Versioning needs precision. The model, code, parameters, data schema, third-party components and user interface can change on different dates. The committee should know which complete system produced the score and whether that combination remains approved.

3. State intended use and prohibited use

Intended use should identify the borrower population, product, geography, decision, horizon, output, user and conditions. Prohibited use should address populations, decisions or circumstances unsupported by evidence.

Using a model beyond its original application creates additional uncertainty. A score validated for portfolio triage may be unsuitable for pricing, covenant setting or decline decisions. Extension requires analysis and approved controls.

Users should see the boundary at the point of decision. A policy hidden in technical documentation provides weak protection against misuse.

Purpose should also define the target and observation window. Probability of default, expected loss, rating transition, liquidity stress and recovery are distinct outputs. A one-year default score should not be described as a complete view of long-term credit quality or transaction value.

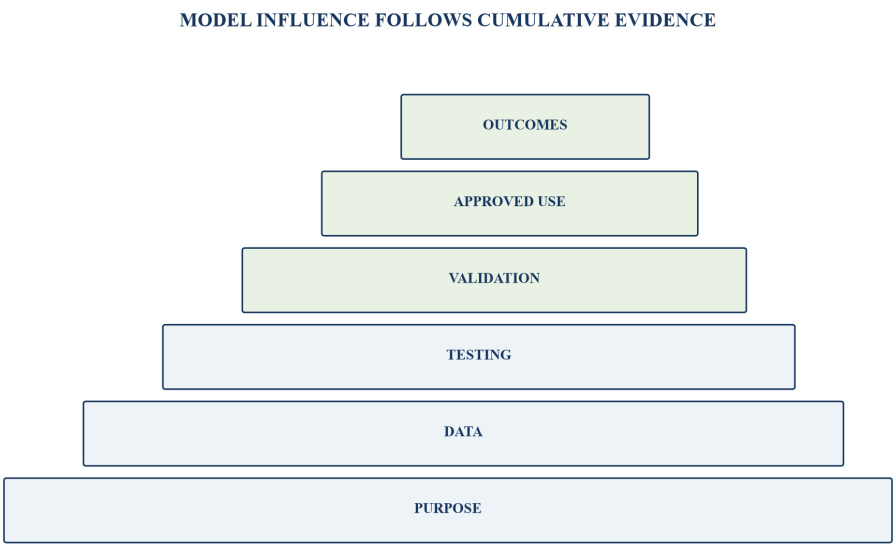
The population statement should describe exclusions and edge cases. Start-ups, project vehicles, financial institutions, public-sector entities or rapidly changing borrowers may behave differently from the development population. Unsupported cases should route to a controlled alternative method.

Prohibited use should be technically and procedurally enforceable where practical. Access, workflow, interface labels and approval rules can prevent a triage score from entering automated approval or pricing. Monitoring should detect attempted or accidental misuse.

4. Build the model evidence ladder

The evidence ladder should progress from registered purpose and controlled data through testing, independent validation, approved use and realised performance. A weakness at a lower stage constrains reliance at higher stages.

Figure 1. Model evidence ladder for credit decisions



Model influence should increase only after evidence and governance gates are satisfied.

Table 1. Minimum committee model-evidence record

Dimension	Required evidence	Control question	Committee use
purpose	decision, population and horizon	is this the approved use	relevance
data	provenance, quality and cut-off	are inputs reliable	confidence
method	design, assumptions and limits	is the approach sound	interpretation
validation	independent tests and findings	is use supported	reliance
performance	stability, calibration and outcomes	does it remain fit	monitoring
override	reason, authority and expiry	is judgement controlled	adjustment
fallback	manual method and reconciliation	can the decision proceed safely	resilience

Every material model output should have a controlled evidence path.

5. Tier model risk by impact

Tiering should consider decision consequence, exposure, model complexity, uncertainty, data sensitivity, user reach, frequency and substitutability. A simple high-impact score can warrant stronger control than a complex low-impact research tool.

The tier determines validation depth, approval authority, monitoring frequency and issue escalation. It should be reviewed when use, portfolio or model characteristics change.

Aggregate model risk also matters. Several models may rely on the same data, macro assumptions or vendor and fail together.

Impact should consider both direct and indirect effects. A model may not approve a facility but can shape which borrowers receive attention, how analysts frame questions or which cases reach committee. Broad influence can create material risk even without formal automation.

Tiering evidence should be documented and challengeable. Complexity, uncertainty and exposure can change over time. A model adopted across new products or geographies may require re-tiering before validation or governance catches up.

The inventory should distinguish active, restricted, under-remediation and retired systems. A retired model can persist in spreadsheets, reports or user habits. Decommissioning needs access removal, archival evidence and confirmation that downstream processes no longer depend on it.

6. Establish authoritative input data

Inputs should have documented sources, ownership, definitions, cut-offs, transformations and quality controls. Missing, stale, disputed and manually adjusted values should remain visible.

Credit data can reflect selection, survival and reporting bias. Development samples may differ from current borrowers or exclude outcomes that matter in stress.

The committee should understand which inputs are observed, estimated, borrower-provided, third-party or model-generated. Material uncertainty should affect permitted reliance.

Source controls should reconcile record counts, totals and key fields through the transformation chain. Manual changes need a reason, evidence, owner and approval. A technically complete feed can still be conceptually wrong when definitions differ from the model design.

Missing-data treatment should be explicit. Imputation can preserve calculation while reducing certainty; default values can create hidden directional bias. The system should show the extent, method and sensitivity of missingness for the individual decision and portfolio.

Outcome data need equal attention. Default, loss, recovery or rating labels can be delayed, revised or affected by policy and intervention. Weak outcome definitions undermine calibration and validation even when input data appear clean.

7. Test representativeness and coverage

The development and validation population should match intended use across borrower type, size, sector, geography, product, credit quality and economic conditions.

Sparse defaults or rapid product change can limit statistical evidence. Alternative outcomes, conservative boundaries and expert review may support a bounded use while limitations remain explicit.

Out-of-distribution cases should trigger restriction or escalation rather than silent extrapolation.

Coverage analysis should compare development, validation and current production populations. Shifts in sector, leverage, size, geography, accounting, product or borrower acquisition channel can weaken relevance. Summary averages should be supported by segment detail.

Selection effects matter. Approved borrowers are observed under lending conditions and interventions; declined applicants may have limited outcomes. A model trained only on past approvals can inherit policy choices and miss risks in a changed strategy.

Stress-period representation should be assessed. A dataset dominated by benign conditions may provide precise but fragile relationships. Scenario analysis and conservative boundaries can supplement evidence while the limitations remain visible.

8. Assess conceptual soundness

Conceptual review examines the relationship between inputs, method, output and credit outcome. It should challenge assumptions, transformations, interactions and the economic rationale for material drivers.

Complexity needs justification through improved decision evidence or performance. Additional parameters can increase instability and make validation difficult.

Alternative models, benchmarks and sensitivity tests help identify dependence on one method. A plausible narrative alone does not demonstrate soundness.

Feature design should reflect information available at the decision date. Leakage from future events, revised accounts or post-decision classifications can inflate historic performance. The validation team should reproduce the timing and availability of every material input.

Interactions require business interpretation. A variable may behave differently across sectors or leverage levels. The committee does not need every mathematical detail, but it needs the assumptions and failure modes that can materially affect the borrower decision.

Benchmarking can include simpler scorecards, expert rules and fundamental analysis. A more complex model should demonstrate an evidence benefit proportionate to its additional uncertainty, validation burden and operating risk.

9. Test discrimination and ranking

Discrimination measures whether the model separates stronger and weaker outcomes. Performance should be assessed out of sample, out of time and by material segment.

A strong portfolio statistic can conceal weak performance for smaller groups. Confidence intervals, sample size and outcome definitions should accompany headline metrics.

Ranking performance does not establish calibrated probability or suitable pricing. The committee should use each metric for its supported purpose.

Metric selection should reflect outcome balance and cost. A portfolio with few defaults can show misleading accuracy when most borrowers are classified as strong. Receiver-operating, precision-recall, lift and rank measures provide different views and should be interpreted with sample size.

Performance by decision band matters. Errors near an approval or review threshold can have greater consequence than errors far from it. Validation should examine boundary cases and whether small input changes create unstable movement.

Comparisons need a stable basis. Changes in borrower mix, lending policy and outcome windows can alter performance independently of the model. Reports should separate these effects where evidence permits.

10. Test calibration and loss estimates

Calibration compares predicted probabilities or losses with realised outcomes. It depends on horizon, definition, observation window and portfolio mix.

Back-testing should show central tendency, tails and segment results. Economic changes can cause temporary or structural deviation.

Overlays or recalibration require evidence, governance and version control. Outputs should be reported before and after material adjustments.

Calibration can be assessed in bands and segments, with attention to tail outcomes. A model can be well calibrated in aggregate while underestimating risk in a concentrated sector or recent vintage. The committee should see material pockets and proposed restrictions.

Loss estimation adds exposure and recovery assumptions. Probability, loss severity and exposure at default may respond differently to stress. Dependencies and double counting should be tested when components are combined.

Recalibration can improve alignment while leaving conceptual weakness unresolved. The approval memo should state whether a change adjusts level, ranking, segmentation or method and

what evidence supports continued use.

11. Examine stability and drift

Stability analysis should cover inputs, population, score distribution, relationships, outcomes and user behaviour. Drift can arise from markets, lending strategy, borrower conduct, data feeds or model changes.

Thresholds need defined owners and responses. A breach can trigger investigation, restriction, recalibration, independent review or withdrawal depending on consequence.

Gradual change matters. Many small updates can accumulate into a material shift without an obvious approval event.

Drift measures should distinguish data quality breaks from genuine population change. A sudden feature shift can reflect a system migration, borrower behaviour or economic stress. Investigation should identify the cause before recalibration conceals it.

Monitoring windows need balance. Short windows provide speed but can be noisy; long windows can delay response. The chosen cadence should reflect portfolio size, outcome frequency, model impact and the availability of leading indicators.

User drift also matters. Analysts may learn to work around a model, change manual inputs or rely on one output more heavily than approved. Workflow and override records can reveal changes in practical use.

12. Validate implementation

Production code, data mappings, transformations, parameters and interfaces should match the approved design. End-to-end tests should reproduce expected outputs.

Adverse cases should include missing data, extreme values, stale feeds, access failure and conflicting sources. User acceptance should test interpretation and workflow.

Change control should prevent unapproved code, prompt, vendor or configuration updates from altering a lending decision.

Implementation testing should compare independent calculations with production output across normal, boundary and adverse cases. Test data should include missing, duplicated, delayed and contradictory records. Results and tolerances need approval before release.

Interfaces should preserve meaning. A probability, grade, confidence indicator and recommendation are different fields. Rounding, colour scales or labels can change user perception and must be tested as part of the system.

Operational controls should verify job completion, data freshness, access, logging and reconciliation. A failed pipeline should stop or flag model-dependent use rather than silently reuse an old score.

13. Design explainability for the decision

Explainability should identify material drivers, direction, sensitivity, uncertainty and limitations relevant to the committee. It should support challenge and action.

Local explanations can vary for similar cases and may not represent causal effects. Global summaries can conceal borrower-specific interactions. Both require careful interpretation.

A generated narrative should link to controlled evidence. Persuasive language must not substitute for validation or credit judgement.

Driver explanations should be stable enough for the intended use. If small, immaterial input changes produce different explanations, users may draw inconsistent conclusions. Validation should test faithfulness, robustness and segment behaviour.

Explanations should avoid causal language unless the method and evidence support it. A feature associated with higher risk does not prove why the borrower is risky or which action will improve the outcome. Fundamental analysis remains necessary.

Committee packs should present uncertainty and counterevidence. A ranked list of drivers can be accompanied by sensitivity, missing data, unsupported segments and known interactions. This supports challenge rather than rhetorical certainty.

14. Preserve human decision rights

The committee determines credit action within delegated authority. The model owner, validator, user, credit officer, legal adviser and control functions have distinct responsibilities.

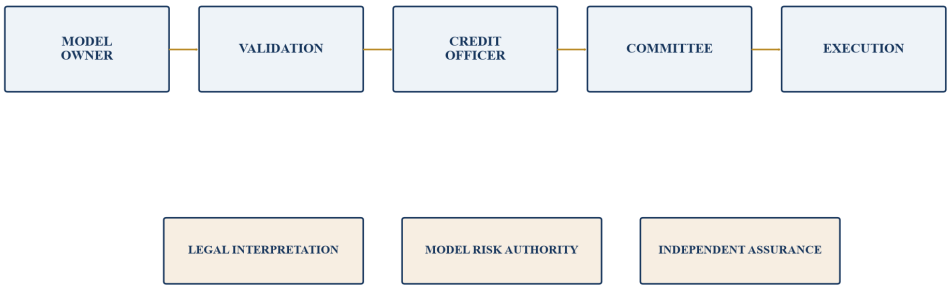
The decision-rights map should show who can approve use, accept findings, impose restrictions, override output, change terms and stop the model.

Model development should not approve its own evidence. Independent validation assesses conceptual soundness, implementation and performance, while business and credit owners decide whether the remaining risk is acceptable for the use.

Legal and compliance functions determine applicable obligations within their mandates. A model score does not interpret a contract, establish discrimination compliance or satisfy disclosure duties. Questions should route to qualified owners.

Emergency authority should be defined before failure. Named roles should be able to restrict or suspend model influence when data, performance, security or use breaches a material boundary. The action and restoration criteria belong in the incident record.

Figure 2. Decision-rights map for model-informed lending



Development, validation, credit judgement and legal interpretation remain distinct.

15. Govern overrides and expert judgement

An override may address data error, evidence outside the model, policy exception or temporary uncertainty. The record should state original output, adjusted treatment, reason, evidence, authority, effective date, expiry and outcome.

Overrides should remain in performance analysis. Repeated patterns can reveal weak inputs, segmentation, calibration or incentives.

Authority should reflect impact. An adjustment affecting ranking differs from one changing approval, limit or pricing.

The institution should distinguish model override from credit decision. A credit officer can approve a different facility structure while leaving the model output unchanged; changing the score itself affects performance history and should require a separate rationale.

Temporary overrides need expiry and review. Conditions can change quickly, and a justified adjustment can become stale. The system should surface expiring items before the next decision or monitoring cycle.

Outcome testing should compare original output, override, final decision and realised result. This allows management to evaluate the model and judgement without treating one favourable outcome as proof.

16. Analyse override patterns

Override analysis should examine direction, magnitude, frequency, user, segment, reason, expiry and realised performance. Both upward and downward adjustments matter.

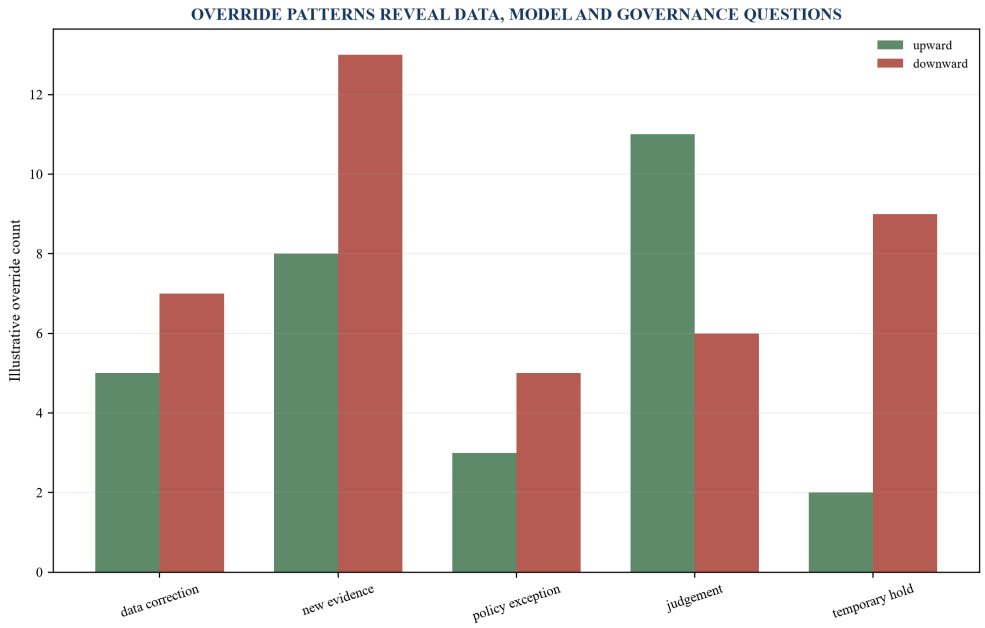
Concentration among users or portfolios may indicate inconsistency or local knowledge. Investigation should distinguish legitimate context from systematic bias.

Magnitude analysis matters. Many small overrides and a few large overrides create different risks. Directional patterns can reveal optimism, conservatism, threshold management or missing risk factors.

Review should examine cases without overrides. A low override rate can reflect an excellent model, weak challenge or workflow friction. User interviews, rejected requests and decision minutes provide context.

Remediation can address data, features, segmentation, calibration, policy, training or authority. The response should match the cause rather than impose a generic override cap.

Figure 3. Illustrative override pattern analysis



Counts and outcomes are hypothetical assumptions for governance design.

17. Control alternative data

Alternative data can improve timeliness or coverage while introducing provenance, consent, bias, stability and explainability risk. The institution should document relevance and permitted use.

Proxy effects require testing. A variable can correlate with protected or inappropriate characteristics even when the field is absent.

Third-party data need diligence, quality monitoring, contractual rights, continuity and exit arrangements.

Relevance should be demonstrated for the credit outcome and timing. Digital activity, transactions, location or behavioural measures can change when platforms, privacy choices or economic conditions change. Stability monitoring should reflect the data-generation process.

Consent and disclosure requirements can vary by jurisdiction and borrower type. The evidence record should identify the legal and policy basis for collection, transformation, sharing and decision use.

Adverse-action or explanation obligations may require accessible reasons. A model should not rely on variables or transformations that prevent the institution from meeting applicable

responsibilities.

18. Govern vendor models

The institution remains responsible for understanding purpose, design, data, limitations, validation and performance. Proprietary restrictions should not prevent appropriate challenge.

Customisation, overlays and local calibration require documentation and testing. Vendor updates should enter controlled change management.

Fallback and data portability matter when the model supports material credit decisions.

Vendor documentation should cover development population, methodology, validation, limitations, updates and incidents to the extent required for responsible use. Marketing claims or certifications do not replace institution-specific testing.

Local implementation can create new risk. Field mapping, data preprocessing, thresholds, overlays and interface design should be validated in the institution's environment and population.

Contracts should support notification, audit or assurance, performance evidence, security, resilience, data access and orderly exit according to service criticality. Concentration in a common vendor should enter aggregate risk reporting.

19. Distinguish generative AI from scoring models

Generative AI can retrieve, summarise or draft material around a credit decision. It may sit outside guidance written for traditional quantitative models, while broader governance, security and accountability remain relevant.

The system-level view should cover prompts, retrieval, foundation models, tools, rules and human review. Each component can change independently.

Generated content should never become an unverified input to approval. Material statements require source linkage and accountable review.

Retrieval should use approved documents with version and access controls. Executed agreements, amendments, financial statements and current management information should be distinguished from drafts and commentary.

Prompt injection, malicious documents, hallucination and sensitive-data leakage require testing and monitoring. External content should be treated as data rather than instruction within the system.

The record should retain prompts, retrieved sources, model version, output and reviewer edits when generative content supports a material decision. Manual fallback should remain available when source accuracy cannot be verified.

20. Define approval conditions

Approval should state permitted use, population, users, limits, data requirements, validation findings, monitoring, overrides, fallback and expiry or review date.

Open issues need owners, deadlines and consequence. A conditional approval should not become indefinite use without closure evidence.

The committee should know which failures suspend model influence and which allow continued bounded use with compensating controls.

Conditions should be testable and tied to consequence. A data remediation item may allow triage while blocking pricing; an unresolved validation finding may require conservative use or independent approval. The memo should state the exact boundary.

Approval duration should reflect uncertainty and change. New models, new populations or rapidly changing data can require shorter review cycles and more intensive monitoring. Renewal should rely on current evidence rather than the original case.

Closure evidence should be independently reviewed where material. A management assertion that an issue is fixed is insufficient without testing, documentation and acceptance by the designated authority.

21. Build the drift dashboard

The dashboard should show population, inputs, score distribution, calibration, discrimination, overrides, issues, data quality and outcomes. It should identify version and monitoring cut-off.

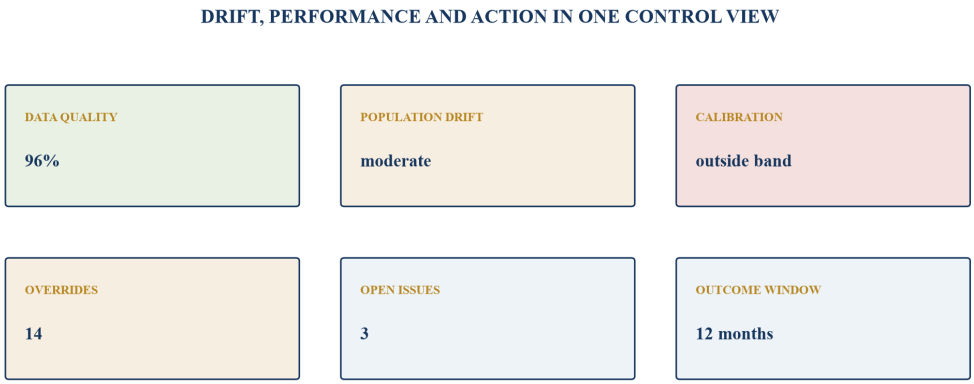
Thresholds should connect to named responses and authority. Traffic lights without consequence can create false comfort.

The dashboard should show both current level and change. A stable but weak calibration measure differs from a sudden deterioration, and the response may differ. Commentary should identify cause, affected decisions and action status.

Segment drill-down is essential. Aggregate green status can conceal a high-impact pocket, while a small weak segment can distort a portfolio average. The committee should see exposure and decision count alongside statistical measures.

Data and model status should appear beside each score used in a live decision. Portfolio monitoring alone can miss a case produced during a feed failure, expired approval or restricted model version.

Figure 4. Illustrative model drift and performance dashboard



Scores, counts and thresholds are hypothetical assumptions.

Table 2. Monitoring threshold and response matrix

Indicator	Evidence	Trigger question	Response
data quality	completeness and reconciliation	are inputs reliable	investigate or restrict
drift	population and input change	is validation still relevant	segment review
calibration	predicted versus realised	are estimates aligned	recalibrate or overlay
discrimination	ranking performance	does separation persist	restrict influence
overrides	pattern and outcome	is judgement controlled	governance review
incidents	failures and misuse	is safe operation possible	suspend or fallback

Actual thresholds require model-specific approval.

22. Monitor data quality as decision evidence

Completeness, validity, uniqueness, timeliness and reconciliation should be assessed by source and segment. Material missingness should affect confidence and use.

Data breaks need an incident record, owner, affected decisions and remediation. Silent imputation can obscure uncertainty.

Quality thresholds should reflect field importance and consequence. A missing immaterial attribute differs from an absent liquidity value or incorrect borrower identity. Controls should prioritise decision-critical data.

Reconciliation should follow the data from source through feature and score. Totals, distributions and individual samples can detect different failure types. Changes in source systems need pre- and post-implementation comparison.

The institution should identify decisions made during a material break and determine whether review, notification or remediation is required. Restoration includes reconciliation with the

corrected output.

Data-quality reporting should distinguish borrower-specific exceptions from systemic defects. One missing document may require case review; a mapping error across a portfolio can invalidate many scores and requires incident governance. Exposure, decision count and time window help determine materiality.

Ownership should extend to derived features. A source system owner may confirm the raw field while the model team owns transformations and joins. Reconciliation and validation should identify where meaning can change and who approves corrections.

Committee users need practical visibility. The pack should state whether the current borrower contains missing, stale or adjusted fields and how those conditions affect the score. Portfolio averages alone do not answer the live decision question.

Periodic assurance should sample the complete data path for decisions across risk bands and products. Reviewers should compare source records, transformations, score inputs, displayed output and committee evidence. Findings need owners, deadlines, acceptance criteria and retesting. This detects silent defects that threshold monitoring can miss and confirms that remediation changed the production decision chain.

Assurance results should reach model-risk governance and credit committees through a tracked record that identifies affected decisions and residual exposure.

23. Measure realised performance

Outcome analysis should compare predictions with defaults, losses, recoveries, rating movement and other approved outcomes over suitable windows.

Actions taken after a score can alter the outcome. Evaluation should document intervention and avoid simplistic attribution.

Outcome windows should match purpose. A short-horizon liquidity model, one-year default estimate and multi-year recovery model require different observation periods and intermediate measures.

Censoring and incomplete outcomes should be handled transparently. Recent vintages may appear strong because adverse events have not matured. Reports should distinguish mature and emerging evidence.

Decision quality should be assessed as well as prediction. Structure, monitoring and conditions can reduce loss even when deterioration occurs. The record should allow analysis of score, judgement, action and result.

Outcome analysis should include adverse cases that the model ranked as strong and benign cases ranked as weak. Case review can reveal missing factors, data errors, unstable relationships, intervention effects or reasonable uncertainty. Findings should feed model and policy improvement.

Benchmark performance should be maintained through time. A model that outperformed a simpler approach during development may lose that advantage as portfolio and economic conditions change. Continued complexity needs current evidence.

The institution should avoid target leakage in outcome review. Revised classifications, post-decision collections or later financial statements must not be treated as information available at the original score date. Reconstructed decision-time data support fair testing.

24. Control overlays and post-model adjustments

Overlays can address risks not captured by the model. They need rationale, method, amount, authority, monitoring and release conditions.

Results should be visible before and after adjustment. Independent challenge should examine material overlays.

Portfolio overlays and borrower-specific adjustments should be separated. A macroeconomic uncertainty may affect a broad population, while new borrower evidence affects one case. Their methodology and release conditions differ.

Overlays should not compensate indefinitely for a model that no longer performs. Recurring or expanding adjustments can indicate the need for recalibration, redevelopment, segmentation or withdrawal.

Committee reporting should avoid double counting. If a model input or scenario already captures a risk, an additional overlay needs a distinct rationale and sensitivity analysis.

25. Manage model limitations

Limitations should state cause, affected population, consequence, mitigation, owner and review date. Materiality determines whether use is restricted.

A generic disclaimer provides weak governance. Users need limitations connected to the current decision.

Limitations can arise from data scarcity, proxy targets, simplified assumptions, unstable relationships, restricted validation or operational dependencies. Each should state how it can bias or destabilise output.

Mitigants should be specific. Additional documents, independent calculation, conservative limits, senior approval or restricted use can reduce risk. A control that does not address the limitation should not support reliance.

Known limitations should enter training and interface design. Users need practical examples of decisions that remain inside and outside the approved boundary.

26. Design manual fallback

Fallback should identify source data, calculation, judgement, review, authority and reconciliation. It should be tested through realistic outages.

Manual work can introduce its own error and capacity risk. The committee should know when delay is safer than unsupported approximation.

Fallback should preserve the distinction between model and judgement. A manual score that imitates the unavailable model without validated inputs can create false assurance. The alternative should have its own documented basis and authority.

Testing should include peak workload and simultaneous failures. Staff, data and review capacity may be sufficient for one case but inadequate during portfolio stress. Prioritisation and escalation routes should be rehearsed.

After restoration, the institution should compare manual and model results, investigate differences and confirm which decision record is authoritative.

27. Protect confidential and personal data

Credit models can use financial, customer, employee and behavioural data. Purpose, access, retention and transfer controls should reflect sensitivity and law.

Data minimisation, secure environments and incident response protect both borrower and lender. Model access should follow least privilege.

Development and validation environments should use controlled data. Copies, extracts and vendor transfers can multiply exposure. Inventory, encryption, logging and disposal should reflect the sensitivity of the source.

Model outputs can themselves be sensitive. Scores, explanations and inferred attributes may affect borrower relationships and decisions. Access and disclosure should follow approved purpose.

Security incidents can undermine data integrity as well as confidentiality. Response should identify affected model versions, decisions and evidence and determine whether re-scoring or committee review is required.

28. Address fairness and proxy risk

Testing should examine whether inputs or outcomes create unexplained disparities across relevant groups. Population and legal context determine the appropriate analysis.

Removing a field does not remove proxy effects. Remediation may require data, design, policy or use changes.

Testing should reflect the decision pathway, including overrides and cut-offs. A statistically balanced score can lead to unequal outcomes through thresholds, data availability or user behaviour.

Observed differences require careful interpretation. Sample size, legitimate risk factors, data quality and portfolio selection can affect results. Qualified legal and compliance review determines applicable obligations.

Monitoring should continue after launch because populations and relationships change. Findings need owners, action and retesting rather than one-time certification.

29. Preserve effective challenge

Challenge requires competence, independence, authority and time. Validators and credit officers should be able to question assumptions, use and findings.

Commercial urgency should not compress review below the approved evidence threshold. Unresolved dissent belongs in the committee record.

Challenge should be informed by access to data, documentation, code or equivalent evidence, test results, issues and users. Proprietary barriers may require additional controls or restricted use.

Independence should be supported by reporting lines, incentives and authority. A validator who identifies a material weakness needs an escalation path capable of restricting use.

Committee members should challenge both model optimism and reflexive rejection. The purpose is to understand supported influence and residual risk rather than seek a predetermined answer.

30. Prepare the approval memo

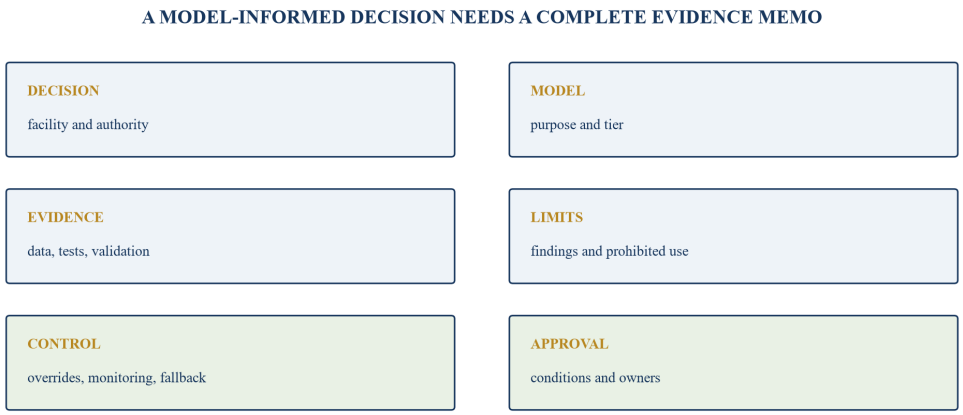
The memo should state decision, model, purpose, population, tier, data, method, validation, performance, limitations, overrides, conditions, fallback and requested authority.

The memo should identify the complete model-system version and effective date. It should state whether current production matches the validated configuration and whether any components or data changed after testing.

Material findings should appear in the main decision section, not only an appendix. Each needs impact, mitigation, owner, deadline and the consequence of non-completion.

The recommendation should explain how much influence is requested. Terms such as support, inform or assist are ambiguous without permitted decisions and prohibitions. The committee should approve a precise boundary.

Figure 5. Credit-committee model approval memo architecture



The memo should connect evidence, limits, authority and monitoring.

Table 3. Approval memo evidence and decision fields

Section	Minimum content	Evidence owner	Decision question
use	decision, population and users	business owner	is use appropriate
validation	scope, tests and findings	independent validator	is evidence sufficient
performance	metrics, segments and outcomes	model owner	does it remain fit
limitations	impact and mitigants	control owner	what reliance is allowed
conditions	action, owner and date	accountable executive	what must be completed
fallback	manual process and testing	operations	can decisions continue

The memo should make reliance and restrictions explicit.

31. Report uncertainty to the committee

The committee should see ranges, sensitivity, sample size, confidence and unresolved evidence. A single score can conceal uncertainty.

Material assumptions should be linked to downside and decision conditions. Precision should match evidence.

Uncertainty can come from sampling, data quality, model specification, economic change and borrower-specific evidence. These sources should not be collapsed into one confidence label when they imply different actions.

Sensitivity analysis should focus on decision relevance. The committee needs to know which plausible input or assumption changes would alter approval, structure or monitoring, and which variations leave the decision robust.

Communication should avoid technical overload while preserving substance. A concise summary can link to detailed validation, data and performance evidence for reviewers who need deeper challenge.

32. Compare model evidence with fundamental credit analysis

Cash flow, leverage, liquidity, structure, collateral, management and market evidence remain central. The score provides an additional lens.

Conflicts between model and fundamental analysis require investigation and recorded judgement. Neither should prevail automatically.

Fundamental analysis can also be biased or inconsistent. The model may reveal a pattern that deserves challenge, while borrower-specific evidence may identify conditions absent from the training data. The committee should test both.

The comparison should use aligned definitions and dates. A model score based on last-quarter accounts cannot be compared directly with a current liquidity forecast without recognising timing and scope.

Structure can change risk after scoring. Collateral, covenants, amortisation, guarantees and conditions may mitigate or create exposures not represented in a borrower-level score. Final terms

require transaction-specific analysis.

33. Run a hypothetical worked example

Consider a hypothetical borrower seeking a USD 40 million facility. A model assigns a 2.1 per cent one-year probability of default and a strong relative rank. Values are illustrative assumptions.

Validation shows acceptable discrimination, modest calibration error and weaker performance in the borrower's fast-growing segment. Data quality is 94 per cent because two operating feeds are incomplete.

The score's principal drivers are historic cash generation, leverage and payment history. It does not capture the pending renewal of a customer representing 28 per cent of revenue. These figures are hypothetical assumptions for framework demonstration.

The model is current and approved for ranking and analyst support. A validation finding requires additional review before use in material limit or pricing decisions for the fast-growing segment.

34. Apply the evidence ladder

The model is approved for triage and credit review in the segment, with a restriction against automated pricing or approval. The score can inform questions and downside analysis.

The committee receives fundamental evidence showing narrowing liquidity and customer concentration. It considers the model score within that wider record.

The evidence ladder shows controlled purpose, adequate core data, sound implementation and independent validation with a material segment limitation. This supports bounded influence rather than full reliance.

The analyst documents the missing operating feeds and tests the score using conservative values. The rank changes modestly, while the fundamental liquidity concern remains. The decision pack preserves both outputs.

35. Test an override

The credit officer proposes a downward risk override because a major customer contract is due for renewal. Evidence includes contract terms, customer correspondence and concentration analysis.

The override has authority, a six-month expiry and a review condition. It remains visible in performance testing.

The proposed treatment does not rewrite the base model. It records a credit judgement alongside the score and states which terms or monitoring recommendations it affects. This preserves performance evidence.

Independent credit challenge confirms the concentration evidence and requests a downside case. The committee sees the original score, adjusted assessment, evidence and sensitivity before deciding.

36. Decide permitted influence

The committee permits the model to support risk ranking and monitoring. It does not permit the score to determine approval, pricing, covenant or limit.

Approval is conditional on verified liquidity data, enhanced reporting and independent review of the segment finding. The credit decision remains a committee judgement.

The minutes state that the model affected question prioritisation and monitoring intensity. Approval and structure were determined from cash flow, customer concentration, downside capacity and the completed evidence package.

If the missing feeds are not remediated by the deadline, the model cannot influence monitoring for this borrower. Manual review and committee escalation become the approved fallback.

Table 4. Hypothetical model-influence decision matrix

Use	Evidence position	Permitted influence	Condition
triage	validated ranking	prioritise review	monitor drift
approval	mixed segment evidence	inform only	human decision
pricing	not validated	prohibited	further analysis
covenant	no causal support	prohibited	fundamental design
monitoring	current data limits	bounded use	data remediation

Values and outcomes are illustrative assumptions.

37. Implement in ninety days

Days One to Thirty can establish inventory, tiering, evidence standards and decision rights. Teams can reproduce one current model decision.

Days Thirty-One to Sixty can test data, validation, overrides, monitoring and memo workflow. Models should run under existing authority.

Days Sixty-One to Ninety can pilot committee packs, fallback and outcome review before approving broader use.

The initial scope should include models with material use and accessible evidence. A bounded pilot can test data lineage, validation summaries, committee interpretation, override recording and conditions without changing delegated authority.

Implementation measures should include pack accuracy, review time, user understanding, unresolved findings, override quality and completion of decisions. Technology delivery alone does not prove model-risk control.

Expansion should add one decision type or population at a time. Each change should confirm data, validation, roles and fallback before broader model influence is approved.

38. Establish operating roles

Model development, validation, credit, finance, legal, data, technology, security, operations and audit need defined responsibilities.

Issue escalation should work across deadlines, staff absence and vendor incidents. Training should cover purpose, limitations and misuse.

The model owner maintains design and performance; independent validation tests evidence; model-risk governance sets standards and restrictions; credit users interpret output within purpose; committees decide facilities; operations preserve controlled execution.

Internal audit provides independent assurance over governance and control according to its mandate. Findings should reach accountable management and the board or relevant committee with remediation evidence.

Role conflicts need mitigation. Business ownership brings useful context, while independent challenge needs sufficient influence to restrict or stop unsupported use.

39. Use outcomes to recalibrate governance

Realised defaults, losses, recoveries, overrides, decisions and incidents should update analysis. Model and human performance should be distinguished.

Changes require testing, approval and version history. Recent benign results should not erase structural limitations.

Outcome analysis should compare original output, overrides, final decisions, conditions and realised results. It should avoid judging a sound decision solely through one later outcome under uncertainty.

Recurring issues can indicate governance weakness beyond the model. Late validation, ignored restrictions, expired conditions or manual workarounds require process remediation.

Recalibration should preserve traceability to prior versions and disclose breaks in comparability. Committees need to know whether apparent improvement reflects performance or a changed method.

40. Make model influence explicit

A committee should be able to state exactly how the model affected the decision. The record should identify evidence considered, reliance allowed, restrictions, overrides and accountable judgement.

AI can improve consistency and focus when its purpose, data, validation and use remain controlled. Model output does not replace cash-flow analysis, contractual interpretation or credit authority.

The final discipline is a closed path from purpose to evidence, validation, decision, outcome and recalibration. This creates learning while preserving accountability.

Boards and senior management should receive aggregate model-risk information proportionate to the institution's use. Concentrations, material findings, restrictions, incidents and remediation show whether risk remains within tolerance.

Borrowers can benefit from consistent evidence and clearer questions when model use is governed. These benefits require separate measurement and should not be used to overstate predictive accuracy or justify unsupported automation.

The institution should retain the ability to challenge and retire the system. A model that no longer fits purpose, data or portfolio should be restricted, redesigned or withdrawn through authorised governance.

Table 5. Board and committee model-risk assurance record

Dimension	Measure	Evidence	Governance use
inventory	coverage, tier and use	model register	scope
validation	findings and closure	independent reports	reliance
performance	drift, calibration and outcomes	monitoring record	fitness
overrides	reason, authority and expiry	decision records	judgement control
incidents	failures, misuse and response	incident log	resilience
decisions	influence, conditions and outcomes	committee minutes	accountability

Assurance should connect model control with decision consequence.

References

[1] Board of Governors of the Federal Reserve System, Revised Guidance on Model Risk Management, SR 26-2, 2026. https://www.federalreserve.gov/supervisionreg/srletters/SR2602.htm

[2] Federal Reserve, Supervisory Guidance on Model Risk Management, 2026. https://www.federalreserve.gov/frrs/guidance/supervisory-guidance-on-model-risk-management.htm

[3] Bank of England Prudential Regulation Authority, SS1/23 Model Risk Management Principles for Banks, 2026. https://www.bankofengland.co.uk/prudential-regulation/publication/2023/may/model-risk-management-principles-for-banks-ss

[4] Central Bank of the UAE, Credit Risk Management Regulation, 2024. https://rulebook.centralbank.ae/en/rulebook/credit-risk-management-regulation

[5] Central Bank of the UAE, Credit Risk Management Standards, 2024. https://rulebook.centralbank.ae/en/rulebook/credit-risk-management-standards

[6] Central Bank of the UAE, Article 13 Credit Risk Models. https://rulebook.centralbank.ae/en/rulebook/article-13-credit-risk-models-0

[7] Central Bank of the UAE, Monitoring and Validation. https://rulebook.centralbank.ae/en/rulebook/211-monitoring-and-validation

[8] Basel Committee on Banking Supervision, Principles for the Management of Credit Risk, 2025. https://www.bis.org/bcbs/publ/d595.pdf

[9] Basel Committee on Banking Supervision, General Credit Risk Management. https://www.bis.org/committees/bcbs/basel-consolidated-guidelines/module/cri/10

[10] Financial Stability Institute, Regulating AI in the Financial Sector, 2024. https://www.bis.org/publications/fsi-insight-63-regulating-ai-financial-sector-recent-developments-and-main-challenges

[11] Bank for International Settlements, Governance of AI Adoption in Central Banks, 2025. https://www.bis.org/publ/othp90.htm

[12] National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework 1.0, 2023. https://doi.org/10.6028/NIST.AI.100-1

[13] National Institute of Standards and Technology, Generative Artificial Intelligence Profile, 2024. https://doi.org/10.6028/NIST.AI.600-1

[14] National Institute of Standards and Technology, Cybersecurity Framework 2.0, 2024. https://doi.org/10.6028/NIST.CSWP.29

[15] Financial Stability Board, The Financial Stability Implications of Artificial Intelligence, 2024. https://www.fsb.org/2024/11/the-financial-stability-implications-of-artificial-intelligence/

[16] Financial Stability Board, Enhancing Third-Party Risk Management and Oversight, 2023. https://www.fsb.org/2023/12/enhancing-third-party-risk-management-and-oversight-a-toolkit-for-financial-institutions-and-financial-authorities/

[17] European Banking Authority, Guidelines on Loan Origination and Monitoring. https://eba.europa.eu/activities/single-rulebook/regulatory-activities/credit-risk/guidelines-loan-origination-and-monitoring

[18] IFRS Foundation, IFRS 9 Financial Instruments. https://www.ifrs.org/issued-standards/list-of-standards/ifrs-9-financial-instruments/

[19] IFRS Foundation, IFRS 7 Financial Instruments: Disclosures. https://www.ifrs.org/issued-standards/list-of-standards/ifrs-7-financial-instruments-disclosures/

[20] International Organization for Standardization, ISO/IEC 42001 Artificial Intelligence Management System. https://www.iso.org/standard/81230.html

[21] International Organization for Standardization, ISO 31000 Risk Management. https://www.iso.org/iso-31000-risk-management.html

[22] UK National Cyber Security Centre, Guidelines for Secure AI System Development, 2023. https://www.ncsc.gov.uk/collection/guidelines-secure-ai-system-development

[23] Committee of Sponsoring Organizations of the Treadway Commission, Enterprise Risk Management Framework. https://www.coso.org/enterprise-risk-management

[24] Office of the Comptroller of the Currency, Rating Credit Risk, Comptroller's Handbook. https://www.occ.treas.gov/publications-and-resources/publications/comptrollers-handbook/files/rating-credit-risk/pub-ch-rating-credit-risk.pdf

[25] Interagency Statement on the Use of Alternative Data in Credit Underwriting, 2019. https://www.federalreserve.gov/supervisionreg/caletters/CA19-11.htm

About the Author

Authored by Chennakeshav Adya, Independent Researcher

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

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

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

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