IPHO-Journal of Advance Research in Business Management and Accounting,
(e-ISSN 3050-886X, p-ISSN 3050-9327)
Volume 01 | Issue 01 | January 2023
Published by: International Publication House Open (IPHO)
DOI: https://doi.org/10.5281/zenodo.22278120
Available online at: https://iphopen.org/
Journal homepage: https://iphopen.org/index.php/bma

GOVERNED DECISION-INTELLIGENCE (GDI)

AN ANALYTICS-CENTRIC REFERENCE ARCHITECTURE, MODEL-RISK & CONFORMANCE FRAMEWORK, AND ASSURANCE MODEL FOR REGULATED PRODUCTION DECISION SYSTEMS

MESBAUL HAQUE SAZU[*]
[*]Independent Researcher

ABSTRACT

Production analytics increasingly sits in the control path of consequential, often regulated decisions - which claim to pay, which applicant to approve or decline, which risk-adjustment code to submit. The model is only one component, and a technically healthy service can still be an indefensible decision-maker when data quality, feature lineage, calibrated uncertainty, explanation, human oversight, and realized outcomes are governed unevenly. This paper specifies Governed Decision-Intelligence (GDI), a framework whose unit of assurance is the decision rather than the model and whose deliverables are validation, conformance, and audit defensibility. GDI composes established instruments - the SR 11-7 model-risk lifecycle, the NIST AI Risk Management Framework, ISO/IEC management and data-quality standards, exact TreeSHAP attribution, split-conformal prediction, post-hoc calibration, and distributional drift statistics - into one testable control structure delivered as four instruments: a seven-layer reference architecture with an explicit governance spine and closed outcome loop; eight normative invariants at RFC 2119 strength with a conformance rubric; a model-risk and validation methodology; and an evidence-and-explanation model that binds every decision to the sources, versions, calibrated confidence, and deterministic explanation required to reconstruct and defend it. The decision-quality target is grounded in a construct model relating analytics use, competency, and decision quality, with testable hypotheses. The framework is portable across regulated domains and is instantiated for risk-adjustment coding, credit underwriting with adverse-action obligations, governed credit-decisioning delivery, payment-integrity analytics, and as a normative engineering standard. GDI governs how a decision is produced and certified; the run-time operation of the system that then acts on the decision is the concern of a companion production-platform architecture, with which GDI is composable.

Keywords –

governed decision intelligence; decision quality; model-risk management; SR 11-7; conformance; validation; calibration; split-conformal prediction; adverse-action explainability; data veracity; evidence; regulated decision systems.

Corresponding Author's: MESBAUL HAQUE SAZU

Copyright: © The Author(s), 2023. Published by **IPHO-Journal of Advance Research in Business Management and Accounting** This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

1 Introduction

1.1 Motivation and gap

Analytics has moved from retrospective reporting into the direct control path of consequential decisions. In that path the analytical model is one component; the trustworthiness of the eventual decision depends on the quality of heterogeneous inputs, the lineage of engineered features, the calibration of predicted probabilities, the stability of the model as populations shift, the clarity of the explanation offered to a reviewer or an affected party, the boundary between automated and human judgment, and the realized outcome that eventually reveals whether the decision was correct. When these concerns are governed unevenly, predictable failures follow: data quality is checked during transformation but is not a precondition for scoring; outputs are logged without the exact feature, policy, and model versions required to reconstruct them; probabilities are consumed as if calibrated when they are not; human review exists without a calibrated escalation rule or a staffing model; drift is monitored without a defined response. The result is a class of systems that pass their unit tests and meet their latency budgets yet cannot be validated, explained, or defended when an examiner, an auditor, or a disputing party asks how a particular decision was reached.

Prior work supplies strong instruments for individual concerns - model-risk lifecycles, AI risk-management frameworks, data-quality standards, attribution methods, uncertainty quantification, drift statistics, reliability engineering. What is missing is an analytics-centric reference architecture that composes these instruments into one control structure, fixes the boundaries between them, states what must be true at each boundary, and makes the whole testable - treating the analytics capability itself, across the descriptive-to-prescriptive progression, as a first-class governed asset with an explicit relationship to decision quality.

1.2 Contribution and scope

The contribution is not the invention of any single instrument; it is their integration into one decision-centric, testable assurance structure with explicit architectural boundaries, normative invariants, metrics, readiness gates, and an evidence schema. GDI is scoped to regulated and high-consequence decision systems - settings under SR 11-7 expectations, adverse-action and fair-lending obligations, risk-adjustment audit exposure such as RADV, or an internal normative engineering standard against which production systems must be certified. It is deliberately not a platform-operations document: it says what must be true for a decision to be produced and defended, not how to run the service that carries the decision out.

Positioning relative to a production-platform architecture

GDI governs the design-time production and certification of a decision: what must be true before the decision is trusted - veracity, lineage, calibration, deterministic explanation where regulated, model-risk validation, and a measured decision-quality target. It does not specify the run-time operation of the system that then acts on the decision across a multi-step workflow (orchestration, action execution, telemetry, closed-loop retraining, operational readiness); that whole-stack concern is the domain of a companion production-platform reference architecture. The two are complementary and composable.

Adopter lane test. An adopter whose artifact is a model-validation report, a decisioning design specification, a delivered-capability assessment, or a normative engineering standard is working in GDI's lane; an adopter whose artifact is a runtime observability system, an automation or control platform, or an operational-readiness review is working in the companion lane.

2 Background and Contributing Instruments

GDI composes established analytical, statistical, governance, and validation practices, attaching each to a specific control boundary rather than presenting it as an isolated best practice. Table 1 names each instrument and the locus at which it operates. The distinction GDI claims is not novelty of any single instrument but their integration into one decision-centric, testable structure: validation frameworks certify a model in isolation, whereas GDI places the model inside a decision and requires veracity, lineage, calibrated confidence, deterministic explanation, oversight, and an outcome loop around it.

Instrument or method	Contribution to GDI	Locus
Layered big-data pipeline; four-V data model	Data-foundation layering; veracity as an explicit gate	Layers 1–4; INV-1
Analytics-capability & decision-quality model	Capability as mediator; usefulness/effectiveness target	Layer 5; §3; INV-3/4
CRISP-DM analytics lifecycle	Descriptive→prescriptive staging discipline	Layer 4; §7
SR 11-7 model-risk lifecycle	Conceptual soundness, outcomes analysis, monitoring, challenge	§7; INV-4/8
NIST AI RMF; ISO/IEC 42001, 23894	Govern/map/measure/manage; AI management-system controls	Governance spine
ISO/IEC 25012, 5259; ISO 8000	Data-quality dimensions for the veracity gate	Layer 3
DAMA-DMBOK; Model Cards; Datasheets	Registry, provenance, lineage; model/dataset documentation	Governance spine
Exact TreeSHAP / Shapley attribution	Deterministic feature attribution with local accuracy	INV-6; §8
Split-conformal prediction	Distribution-free coverage for confidence gating	INV-7; §7.3
Platt / isotonic / temperature scaling	Post-hoc probability calibration	§7.2
PSI, KS, Jensen-Shannon; Isolation Forest	Distributional drift and anomaly screening	Layer 3; §7.4
Adverse-impact ratio (four-fifths rule)	Portfolio-level fairness monitoring	Spine; §7.4
ISO/IEC/IEEE 42010; ATAM; 29148; RFC 2119	Architecture views; quality-attribute scenarios; requirement strength	§4, §5
Google SRE; distributed tracing (OpenTelemetry)	SLOs, error budgets, readiness; one record per decision	§7; Layer 1

Table 1 · Contributing instruments and their locus within GDI.

3 Theoretical Foundation and Design Commitments

Because GDI targets decision quality rather than model accuracy, it requires an explicit account of what produces decision quality. GDI adopts a three-construct model as its theoretical spine: analytics use raises decision quality both directly and through the mediating construct analytics competency, controlling for data veracity, domain, and model class. Decision quality is operationalized as usefulness (reliability, precision, and accuracy of acted-upon outputs) and effectiveness (time and cost to reach the decision); acceptance and override rates are leading trust indicators.

Hypotheses (adopter-testable on its own telemetry). H1 - greater analytics use is associated with higher decision quality, controlling for veracity and domain. H2 - greater use is associated with higher competency. H3 - greater competency is associated with higher decision quality. H4 - competency partially mediates the effect of use on quality; the indirect path is positive and significant while the direct path remains non-zero. H1 and the veracity control motivate the veracity gate (Layer 3); H2–H4 motivate treating competency as a distinct, funded service whose effect is estimated causally rather than assumed.

Seven design commitments follow from this theory and are made concrete by the invariants of §6:

- **Veracity before analytics.** Data is untrusted until it passes a defined quality gate; sub-threshold data is quarantined, not scored.
- **Named decision before action.** Every acted-upon output is attributable to an explicit, evaluable decision; workflow completion is never a successful outcome.
- **Decision quality as the validation target.** The object certified is the usefulness and defensibility of acted-upon outputs; model discrimination is necessary but insufficient.
- **Evidence binding at generation time.** Reconstruction begins with a record written as the decision is produced, not after-the-fact investigation of mutable logs.
- **Deterministic explanation where regulated.** Where adverse-action or audit obligations apply, attribution must be exact, reproducible, and persisted.
- **Confidence-gated oversight.** Calibrated uncertainty and remit define when automation hands a decision to a human role; escalation is capacity-planned.

- **Governance before scale; close the loop.** Provenance, lineage, access, and monitoring are architecture inputs, not retrofits; realized outcomes return as labels, regression cases, and revalidation triggers.

4 Design Requirements and Formal Preliminaries

Requirements are written to be verifiable in the sense of ISO/IEC/IEEE 29148 and carry RFC 2119 strength. **R1** the data foundation, including a veracity gate, MUST precede analytics. **R2** every automated action MUST trace to an explicit, named decision. **R3** the target metric MUST be decision quality, not model score alone. **R4** analytics competency MUST be a distinct, reusable capability. **R5** each output MUST be bound at generation time to source, feature, model, and policy versions, and the binding MUST be queryable. **R6** explanation depth MUST scale to stakes and MUST be deterministic where regulated. **R7** human review MUST be routed by remit and gated by calibrated confidence. **R8** governance MUST precede scale, and realized outcomes MUST close the loop.

Formal objects. An input record x is scored by a versioned model m to produce a decision $d \in \{approve, decline, refer, submit, hold\}$. Evidence E(d) is the source records, features, versions, policy, and context sufficient to reconstruct d. A veracity gate $V(x) \mapsto \{pass, quarantine\}$ is a pre-analytics predicate; only V(x)=pass reaches analytics. A confidence gate $G(x)=[w(x) \leq \tau]$ tests the split-conformal interval width w(x) against a governed threshold τ. Decision quality DQ = g(U, E) combines usefulness U with effectiveness E. These definitions make the invariants of §6 checkable with implementation evidence.

5 Instrument I - The Reference Architecture

The architecture places seven ordered layers on the forward path from data to action, a governance spine spanning all layers, and a feedback loop returning realized outcomes to analytics and decision support (Figure 1). Layer roles and ordering are fixed; component choices remain open. Each layer exposes a defined analytical responsibility, an input contract, and an evidence contract, so that a veracity failure is not hidden inside a model score, a decision is not inferred from workflow completion, a human override is not lost in a downstream ticket, and an audit does not depend on mutable logs alone.

Figure 1 · GDI seven-layer reference architecture, governance spine, and closed outcome loop

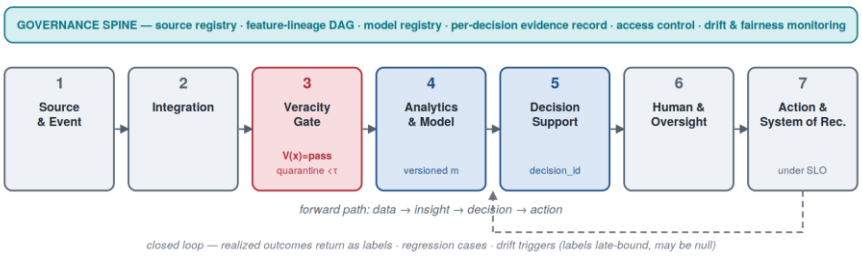


Figure 1 · The seven-layer forward path, the governance spine, and the closed outcome loop.

#	Layer	Analytical responsibility	Evidence / interface contract
1	Source & Event	Capture structured, unstructured, and event data with origin metadata	One record per execution; origin & capture-time metadata persisted
2	Integration	Normalize; reconcile identity by record linkage; expose decision-ready data	Entity keys resolved (Fellegi-Sunter); freshness stamped
3	Veracity Gate	Validate quality dimensions; score source confidence; quarantine below threshold	V(x)=pass required to proceed; quarantine reason recorded
4	Analytics & Model	Descriptive→diagnostic→predictive→prescriptive; versioned models	model_version & feature_set_version bound to each output
5	Decision Support	Bind insight to a named, evaluable decision via a competency service	Every output references an explicit decision_id
6	Human & Oversight	Route by remit; confidence-gated escalation; capture accept/override	Escalation reason and reviewer identity under audit
7	Action & System of Record	Execute under SLO; write realized outcomes back	SLO/error budget enforced; outcome join key emitted

***Table 2** · Layer contracts. Each layer is an assurance boundary with an explicit evidence contract.*

Layer 4 is where the analytics value ladder lives. Descriptive and diagnostic analytics establish baseline distributions and drivers that later monitoring compares against; predictive analytics fits versioned models under a deterministic scoring path; prescriptive analytics converts calibrated predictions into decisions through optimization, uplift, or policy rules. The layer is governed by three standing requirements: models are versioned and registered with a model card; the scoring path is deterministic and reproducible; and the transition to each higher rung is gated on the evidence that rung requires. The governance spine is a plane, not a layer - source registry, feature-lineage DAG, model registry, per-decision evidence record, access control, and drift and fairness monitoring - built before scale because retroactive governance cannot recreate the state of a decision after its context has changed.

Figure 2 · The governed analytics pipeline — veracity gate and confidence gate as load-bearing controls

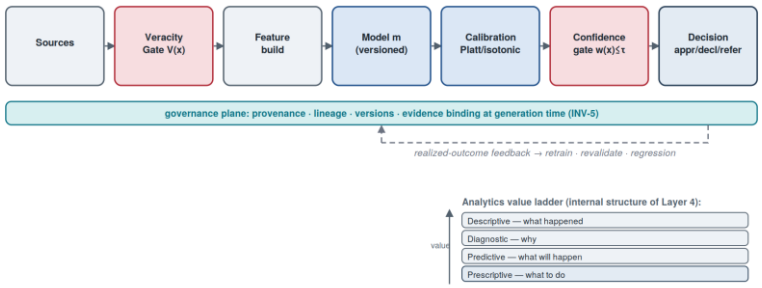


Figure 2 · The governed analytics pipeline; the veracity gate and confidence gate are load-bearing controls, over a governance plane and under a feedback loop.

6 Instrument II - The Eight Invariants

The normative core is eight invariants at RFC 2119 strength, answerable with implementation evidence. A MUST is load-bearing: a production system that fails a MUST is non-conformant. The invariants are what let an architecture diagram be distinguished from a system that merely resembles it, and they map to specific architectural loci (Figure 3).

Ref	Invariant	Requirement
INV-1	Data foundation before analytics (MUST)	Source registration, integration, and a veracity gate precede modeling; sub-threshold data is quarantined, not scored.
INV-2	Decisions before actions (MUST)	Every automated action traces to an explicit, named decision; a completed workflow is never a successful outcome.
INV-3	Judge decision quality (MUST)	The target metric is decision quality - usefulness and defensibility; acceptance and override are leading indicators.
INV-4	Validation independence (MUST)	Validation is performed against a pre-registered protocol by a function independent of development, with effective challenge.
INV-5	Bind evidence at generation time (MUST)	Each output is bound to source, feature, model, and policy versions as produced; the binding is queryable.
INV-6	Deterministic explanation where regulated (SHOULD/MUST)	Explanation depth scales to stakes; where adverse-action/audit obligations apply, attribution MUST be deterministic and persisted.
INV-7	Route human judgment by remit (MUST)	Decisions beyond remit escalate to the correct role; escalation is confidence-gated and capacity-planned.
INV-8	Govern first; close the loop (MUST)	Provenance, lineage, and monitoring precede scale; realized outcomes return as labels, regression cases, and drift triggers.

Table 3 · The eight invariants. Adoptable directly as the normative clauses of an engineering standard.

Figure 3 · The eight invariants mapped to architectural loci (adoptable as a conformance checklist)

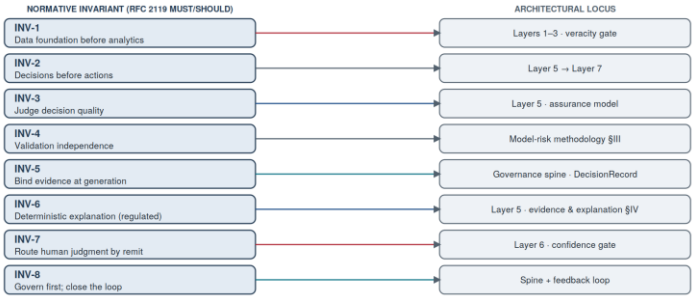


Figure 3 · The eight invariants mapped to architectural loci.

Conformance rubric
Conformant. Every MUST met; every SHOULD met or covered by a recorded, governed deviation. **Partial.** A MUST met unevenly, or a SHOULD unmet without a recorded deviation. **Non-conformant.** Any MUST unmet.
Deviations are legitimate when governed and recorded; the exception mechanism exists so deviations are reviewed rather than silently worked around. A normative engineering standard can adopt these invariants as its clauses and assess production systems clause-by-clause, granting exceptions through a documented deviation process and carrying defects into the next revision.

7 Instrument III - Model-Risk and Validation Methodology

The third instrument fixes how a decision system is validated and monitored so a team can demonstrate, on demand and to an independent challenger, that the system is conceptually sound, calibrated, stable, and fair. It follows the SR 11-7 lifecycle and makes each stage concrete.

7.1 Conceptual soundness and pre-registration

Validation is performed by a function organizationally separate from development, against a protocol fixed before results are observed, with acceptance thresholds set in advance. Conceptual soundness examines the data foundation, feature construction, and modeling choices for fitness to the stated use, and records the boundary of that use - for example, coder-assist with mandatory human adjudication rather than autonomous submission. A pre-registered protocol with Must-criteria and a bounded validation opinion (validated; validated with conditions; not validated) is the artifact an independent model-validation report produces; a single Fail on a Must-criterion withholds validation.

7.2 Calibration

Because a model can rank cases well yet emit miscalibrated probabilities that drive unsafe thresholds, calibration is a first-class control. Probabilities are calibrated post-hoc with Platt scaling, isotonic regression, or temperature scaling and audited with reliability diagrams and the expected calibration error, $ECE = \Sigma_m (|B_m|/n)\cdot|acc(B_m) - conf(B_m)|$ over M bins. A reference pattern of $ECE \leq 0.05$ is a governed default for the task and population, not a universal constant.

7.3 Distribution-free confidence and gated oversight

Predictive uncertainty is quantified with split-conformal prediction, which yields distribution-free marginal coverage $P(Y \in C(x)) \geq 1 - \alpha$ with quantile level $\hat{q} = \lceil (n+1)(1-\alpha) \rceil / n$. The interval width w(x) is the uncertainty signal consumed by the gate $G(x)=[w(x) \leq \tau]$; τ is itself monitored, and review capacity is planned at the same time as τ so escalation rate becomes a direct staffing input rather than a surprise (Figure 4).

Figure 4 · Confidence-gated oversight — a routing contract between analytics and human review

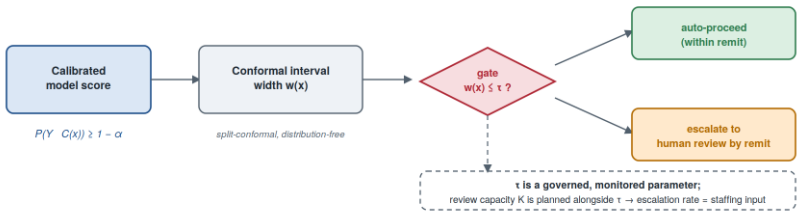


Figure 4 · Confidence-gated oversight as a routing contract between analytics and human review.

7.4 Drift, fairness, and outcomes analysis

Distributional drift is monitored with the population stability index, $PSI = \Sigma_i (a_i - e_i)\cdot\ln(a_i/e_i)$, with the Kolmogorov–Smirnov statistic and Jensen–Shannon divergence. Drift is an operational signal with a defined response: a breach of a governed threshold (a reference pattern near PSI 0.20) opens a retraining or revalidation action rather than a passive alert. Portfolio fairness is monitored with the adverse-impact ratio, AIR = sel_rate(group)/sel_rate(reference), flagged for documented review below the four-fifths threshold. Outcomes analysis compares decisions against an adjudicated gold standard, with inter-rater tie-break where labels are human-generated.

Dimension	Metrics and threshold patterns	Basis
Technical	Latency p50/p95/p99; availability vs SLO; error rate; error budget consumed	SRE
Model	Accuracy/precision/recall/F1; AUC, AUC-PR; ECE ≤ 0.05; regression vs baseline	Metric-selection
Risk	PSI (≈0.20), KS, Jensen-Shannon; anomaly score; adverse-impact ratio	Drift/fairness
Decision	Usefulness & effectiveness; acceptance/override; human-review rate; realized outcome	Decision-quality model

Table 4 · Evaluation dimensions, metrics, and reference thresholds.

8 Instrument IV - Evidence and Explanation

The fourth instrument makes assurance a query rather than an investigation. Every decision persists a single record binding its inputs, versions, calibrated confidence, deterministic explanation, and - late-bound - its outcome.

```
DecisionRecord (generalized evidence schema) {
 decision_id   string
 input_ref   SourceRef[] // source records + registry version
 feature_set_version string
 model_version  string
 policy_version  string
 prediction   float
 confidence   float   // conformal interval width w(x)
 explanation   Factor[]  // deterministic, ranked (TreeSHAP)
 decision   enum { approve, decline, refer, submit, hold }
 outcome    Outcome?  // late-bound; may be null
 audit_reference  string}
```

Framework 1 · The governed DecisionRecord - adoptable as an evidence schema across regulated domains.

Feature attribution uses exact TreeSHAP, which computes Shapley values for tree ensembles in low-order polynomial time and satisfies local accuracy, so contributions sum to the model output offset from the base value: $f(x) = \varphi_0 + \Sigma_i \varphi_i$. Determinism matters because a regulated reason must be reproducible and must match the reason issued to the affected party; non-deterministic reductions such as t-SNE are excluded from the scoring path, and exact attributions are cached where latency requires. A validation gate requires that a sampled historical decision reconstruct end-to-end from recorded versions and that deterministic explanations match the reasons issued: reconstruct(E(d)) = d. This turns an audit or an adverse-action dispute from a log-archaeology exercise into a record lookup.

9 Adoption and Maturity Model

The framework front-loads the governance spine because lineage, evidence, source governance, and monitoring become more expensive and incomplete when reconstructed after scale. Each stage has a measurable exit gate.

Stage	Ships	Gate to advance
0 · Governed skeleton	Source registry & lineage-capturing pipeline; evidence-record write path	Registry complete for in-scope sources; a decision reconstructs end-to-end
1 · Foundation & rules	Veracity gate; deterministic checks; baseline dashboards; decision-ready records	Gate blocks sub-threshold data; lineage queryable
2 · Analytics & evaluation	Versioned models; task metrics; per-decision evaluation; regression	Reproducibility; models validated before promotion
3 · Oversight	Conformal-gated review; role routing; explanation scaled to stakes	Escalation threshold planned with staffing; determinism validated where regulated
4 · Loop & decision quality	Outcome capture; decision-quality evaluation; failures→regression; drift→retrain	Loop closed on realized outcomes; decision quality reported on a cadence

Table 5 · Adoption stages and exit gates (adoptable as a delivery-phasing plan).

10 Instantiation Across Regulated Domains

GDI is one framework with several regulated instantiations. Each leads with a different subset of the four instruments - the natural citation surface for an adopter in that domain (Figure 5).

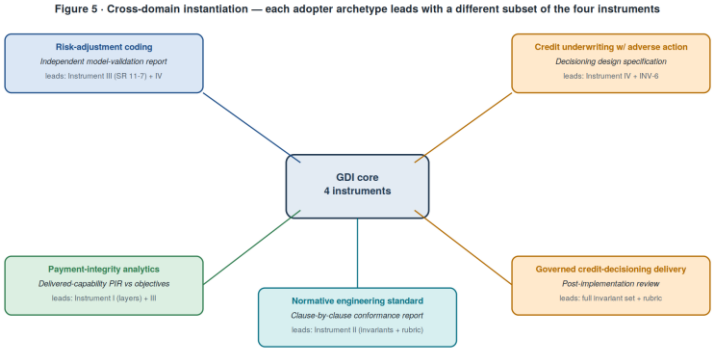


Figure 5 · Cross-domain instantiation; each adopter archetype leads with a different subset of the instruments.

- **Risk-adjustment coding.** An independent validation of a coding model instantiates Instrument III end-to-end: SR 11-7 conceptual soundness in a clinical-coding setting, calibration and discrimination on a pre-registered protocol, subgroup fairness across payer cohorts, and a bounded validation opinion (e.g., coder-assist with mandatory adjudication, autonomous submission prohibited). Representativeness and lineage are first-order validation questions. Cites the validation methodology (§7) and the evidence model (§8).
- **Credit underwriting with adverse action.** A decisioning platform returning approve/decline/refer instantiates the evidence-and-explanation model and INV-6: governed alternative-data lineage, decision traceability reconstructable independently of the originating partner, and factor-accurate, deterministic adverse-action reasons bound at generation time, with confidence gating routing thin-file or uncertain applications to review. Cites §7.3, §8, and the invariants.
- **Governed credit-decisioning delivery.** A post-implementation review of a governance-first decisioning capability instantiates the full invariant set: registry and lineage first, deterministic adverse-action notices as a by-product of the flow, an immutable audit record, and reviewer capacity planned alongside the confidence gate. Cites the conformance rubric and the DecisionRecord.
- **Payment-integrity analytics.** A delivered payment-integrity capability assessed against its original objectives instantiates the architecture and the outcome loop: consolidated sources with end-to-end lineage, models cleared against an agreed precision target on validation and back-testing, an explanation and recommended action shipped with every scored case, and a quantified benefit case. Cites the layer contracts and the validation methodology.
- **Normative engineering standard.** An internal standard for regulated analytics and AI systems adopts the eight invariants as its normative clauses and the conformance rubric as its assessment method, granting recorded exceptions through a deviation process and carrying defects into the next ballot. Cites Instrument II directly.

11 Evaluation

As a reference framework, GDI is evaluated analytically against its own requirements, illustratively through a worked analytics example, and empirically through a protocol testing the §3 hypotheses. Each requirement R1–R8 maps to an instrument feature and a confirming measure (R1→veracity gate and quarantine rate; R3→decision-quality metric and pre/post deltas; R5→evidence record and reconstruction test; R7→confidence gate and correct-routing rate), and no requirement lacks a feature and a measure. The illustrative comparison contrasts an ungoverned baseline with the governed pipeline of Figure 2; the figures show the shape of the evaluation, not empirical results.

Metric	Baseline	Governed	Interpretation
AUC-PR (imbalanced)	0.71	0.74	Comparable ranking power
Expected calibration error	0.14	0.03	Calibrated after isotonic step
PSI at 90 days	0.31 (undetected)	0.31 (flagged, retrained)	Drift converted to a triggered action
Adverse-impact ratio	0.68 (unmeasured)	0.68 (flagged for review)	Fairness surfaced, not hidden
Time-to-trace a decision	hours (log dig)	seconds (record lookup)	Evidence binding pays off at audit
Decision quality (usefulness)	**unchanged**	**higher**	**Acted-upon precision up; overrides down**

Table 6 · Illustrative evaluation - the governed pipeline converts latent risks into surfaced, actioned, auditable events.

The point is not that the governed model ranks better; its ranking power is comparable. The governed pipeline converts latent analytical risks - miscalibration, undetected drift, disparate impact, unreconstructable decisions - into surfaced, actioned, and auditable events, which is what raises decision quality and defensibility. The strongest empirical test is a greenfield implementation in which the framework is specified before the measurement window, a pre-implementation baseline is recorded, the invariants are implemented under explicit exit criteria, and post-implementation decision quality is measured alongside technical, model, and risk metrics; independent review is retained for regulated applications.

12 Threats to Validity and Conclusion

- **Framework, not a controlled causal study.** GDI specifies an architecture and an evaluation method; it does not by itself prove a fixed improvement.
- **Composed methods are not individually novel.** The contribution is integration and normative specification, not the invention of TreeSHAP, conformal prediction, or the ISO standards.
- **Thresholds are governed defaults, not constants.** ECE ≤ 0.05 and PSI ≈ 0.20 are reference patterns to be set for the task and population.
- **Outcome ownership may be external.** Where downstream systems do not expose timely, complete labels, closed-loop learning is limited.

Governed Decision-Intelligence makes the decision, not the model, the unit of assurance, and treats validation, conformance, and audit defensibility as the objects it delivers. Its four instruments - a governed reference architecture, eight normative invariants with a conformance rubric, a model-risk and validation methodology, and an evidence-and-explanation model - compose established controls into one testable structure for regulated decision systems. GDI certifies how each decision is produced and defended; the companion production-platform architecture operates the system that carries it out. The recommended next step is empirical instantiation under controlled measurement.

References

1. Board of Governors of the Federal Reserve System. Supervisory Guidance on Model Risk Management (SR 11-7).
2. National Institute of Standards and Technology. AI Risk Management Framework 1.0.
3. ISO/IEC 42001, Artificial Intelligence Management System; ISO/IEC 23894, AI Risk Management.
4. ISO/IEC 25012 and ISO/IEC 5259 data-quality models; ISO 8000 data quality.
5. ISO/IEC/IEEE 29148, Requirements Engineering; RFC 2119, Key Words for Requirement Levels.
6. Lundberg, S. and Lee, S. A Unified Approach to Interpreting Model Predictions (SHAP); exact TreeSHAP for tree ensembles.
7. Vovk, V., Gammerman, A., and Shafer, G. Algorithmic Learning in a Random World; split-conformal prediction.
8. Platt, J. Probabilistic Outputs for SVMs; Zadrozny and Elkan, isotonic calibration; Guo et al., temperature scaling.
9. Naeini, M. P. et al. Obtaining Well-Calibrated Probabilities (expected calibration error).
10. Population Stability Index; Kolmogorov–Smirnov statistic; Jensen–Shannon divergence for distributional drift.
11. Equal Credit Opportunity Act / Regulation B; adverse-action notice requirements; adverse-impact ratio and the four-fifths rule.
12. Mitchell, M. et al. Model Cards for Model Reporting; Gebru, T. et al. Datasheets for Datasets.
13. Klein, G. and others; analytics-capability and decision-quality construct model (analytics use, competency, decision quality).
14. Kimball / CRISP-DM: the descriptive-to-prescriptive analytics lifecycle; ISO/IEC/IEEE 42010 and ATAM for architecture description and evaluation.

Appendix A - Notation and Conformance Checklist

Symbol	Meaning
x, m, d	Input record; versioned model; explicit, evaluable decision
E(d)	Evidence sufficient to reconstruct d: sources, features, versions, policy, context
V(x)	Veracity predicate; only V(x)=pass reaches analytics
w(x), τ, G(x)	Conformal interval width; governed threshold; confidence gate G(x)=[w(x) ≤ τ]
φ_i	Shapley attribution of feature i; contributions sum to f(x) − base value
U, E, DQ	Usefulness; effectiveness; decision quality DQ = g(U, E)
α	Conformal miscoverage level; coverage ≥ 1 − α

Table A1 · Notation.

Ref	Required evidence for a Conformant verdict
INV-1	Veracity gate present; quarantine rate reported; no sub-threshold data scored
INV-2	Every action links to a decision_id; no action without a named decision
INV-3	Decision-quality metric defined and reported with pre/post deltas
INV-4	Independent validation against a pre-registered protocol; effective challenge recorded
INV-5	Sampled decision reconstructs end-to-end from recorded versions
INV-6	Explanation depth scales to stakes; determinism validated where regulated
INV-7	Confidence gate active; correct-routing rate and staffing model reported
INV-8	Governance spine precedes scale; realized outcomes close the loop on a cadence

Table A2 · Conformance checklist. All MUST items are required for a Conformant verdict.