



AI Governance Meets the External Auditor

Building a Framework for AI Assurance



Welcome to Leading AI Governance – Session 10

AIGOV LEADING AI Governance

Leading AI Governance is an executive webinar series tackling the most urgent challenges in responsible AI. Each month, industry leader Kelle O’Neal cuts through hype to deliver clear, practical frameworks on oversight, risk, regulation, and enterprise-scale governance through a live, candid conversation with a guest executive.

Leading AI Governance Subject Matter Expert:



Kelle O’Neal

Founder and CEO,
First San Francisco Partners

Webinar Title	Date
Cutting Through the Noise: What AI Governance Leaders Really Need to Know	1/6/2026
How to Implement AI Governance: Lessons from Enterprise Leaders	2/3/2026
AI Governance vs. Data Governance Strategic Alignment Without Redundancy	3/3/2026
Human Oversight in AI: Designing Accountability for High-Stakes Decisions	4/7/2026
Governance in the Wild: Managing Shadow AI and Decentralized Models	5/5/2026
Governance for Multimodal AI: Text, Vision, Voice, and Beyond	6/2/2026
Synthetic Truth: Governing Generative AI in High-Stakes Domains	7/7/2026
AI Governance Meets Cybersecurity: Aligning Trust, Safety, and Resilience	8/4/2026
The Right to Explanation: Meeting Regulatory Demands for Interpretable AI	9/1/2026
Building a Framework for AI Assurance	10/6/2026
The Future of AI Governance: Forecasting the Next Five Years	11/3/2026
Scaling AI Governance: Enterprise Playbooks for Data and IT Leaders	12/1/2026

Hosted on the first Tuesday of each month

Can you truly audit a neural network?

Not directly. So we build a broader framework of assurance around the entire AI system.

Building a Framework for AI Assurance

December 1938

McKesson & Robbins, Inc. – The Client

A large, respected drug company, audited for over a decade by one of the most prominent firms in the country, turns out to have about \$20 million in assets that never existed: inventory in Canadian warehouses that weren't there and receivables from customers who weren't real.

December 1938

Price, Waterhouse & Co. – The Auditing Firm

The auditors examined the paperwork management handed them. Observing inventory and confirming receivables directly with customers weren't required practice at the time. The audit met the standards of its day, and those standards were what failed.

1939 - 1940

The New Rules

In 1939 the profession began requiring auditors to physically observe inventory and confirm receivables with third parties.

Reforms also moved auditor selection away from management and toward the board and shareholders.

In 1940 the SEC's report on the case urged that auditors answer to the board and shareholders rather than management.

October 2026

Where AI assurance is today

In 1938, auditing practice was ahead of any agreed standard. AI assurance is in a similar position now. Methods, certifications and auditors exist, but nobody has settled what an AI audit has to examine.

The 1939 reforms came down to two questions:

- Did anyone examine the actual thing?
- Who checked the people who produced the evidence?

You'll be on both sides of this. You'll build the evidence for your own decisions, and you'll be handed someone else's evidence and asked to trust it. The two questions work from either seat.

California starts with the Auditor

September 9, 2026: Governor Newsom signed SB 813 and AB 1405.

SB 813: a voluntary state designation for independent verification organizations. Criteria due January 1, 2028.

AB 1405: an AI Auditor Registry by January 1, 2029, for audits that state law requires.

Neither law requires an organization to be audited.

Bottom line: Independence is being defined. What an audit has to examine is still open.

Where the Responsibility Lives

Governance: The decisions about how much authority AI is given, who approves that authority, the limits it operates within, and how it's measured. It is set before the system acts and revisited as risk changes.

Assurance: The linked set of activities across an AI system's lifecycle that produce the evidence someone independent would need to have justified confidence that the system is doing what was claimed.

Audit: An independent, formal examination, at a point in time, of whether the system conformed to what governance approved, using the evidence assurance produced.

Bottom line: Governance decides what AI may do. Assurance produces the evidence that it did. Audit independently tests that evidence.

Three layers of AI Assurance. Today we focus on one

Layer	The assurance question	Evidence that answers it
Tool	Is the inventory complete?	Detection coverage. Inventory reconciled to network, identity and procurement records. Vendor change notices.
Data	Do the controls operate?	Classification, boundary and DLP logs. Vendor data-use and retention terms. Logged exceptions.
Decision	Does it match what was approved?	The Decision Record, the log and the human action, compared.

The decision layer is where AI changes most what has to be recorded, and where practice is least settled.

The Assurance Chain



Highlighted: the three artifacts this talk follows

A chain, not a checklist: each link consumes what the one before produced. The evidence is housed in **artifacts**, and the artifacts **contain the decisions they record**.

What this talk covers, and why these three

Question

- Should this exist, and how much authority did AI get?
- Artifact
 - Approved use case

Question

- What was the system specified to do?
- Artifact
 - Decision Record

Question

- Did it do that?
- Artifact
 - Audit evidence

WHY THESE THREE ARE THE MOST CRITICAL

They sit on the critical path

No use case, nothing to specify. No record, no criteria to test. No evidence, nothing to attest.

They are where AI changes what must be recorded

The authority conceded, behavior a system can enforce, and proof of conformance. Testing, fairness testing, security and data quality build on disciplines with owners.

They are where industry guidance stops

Certificates like ISO/IEC 42001 and SOC 2 cover the organization. California is creating an auditor registry before any law defines an AI audit.

Also in the chain, owned by established functions: data quality, security, and model testing and validation.

Risk is what AI assurance serves

Before Deployment

- Risk is speculative. Decide what to accept and what to design out.

After Deployment

- Risk is realized. Evidence shows whether the decisions held.



Risk is a property of the decision process, not only the outcome.

Risk enters when we approve the use case

At approval we decide, **decision by decision**, how much authority we concede to AI.



FOR EACH DECISION IN THE USE CASE, ANSWER



From September - Internal Checks Become Visible – External Consistency

Worked example: the credit decline

The same list now has to do more than trace. It has to produce wording specific enough for the applicant to act on, and consistent enough that two applicants with the same underlying issue get the same reason.

Example: *"Debt-to-income ratio: 42%. Approval threshold: 36%."* - not *"unable to approve your request at this time."* Both could technically sit on the approved list but only the first meets this bar.

The check is now external. Not "can the Decision Steward trace this", it's "can anyone reading this letter, with no inside knowledge, tell what to fix and trust it's not arbitrary."



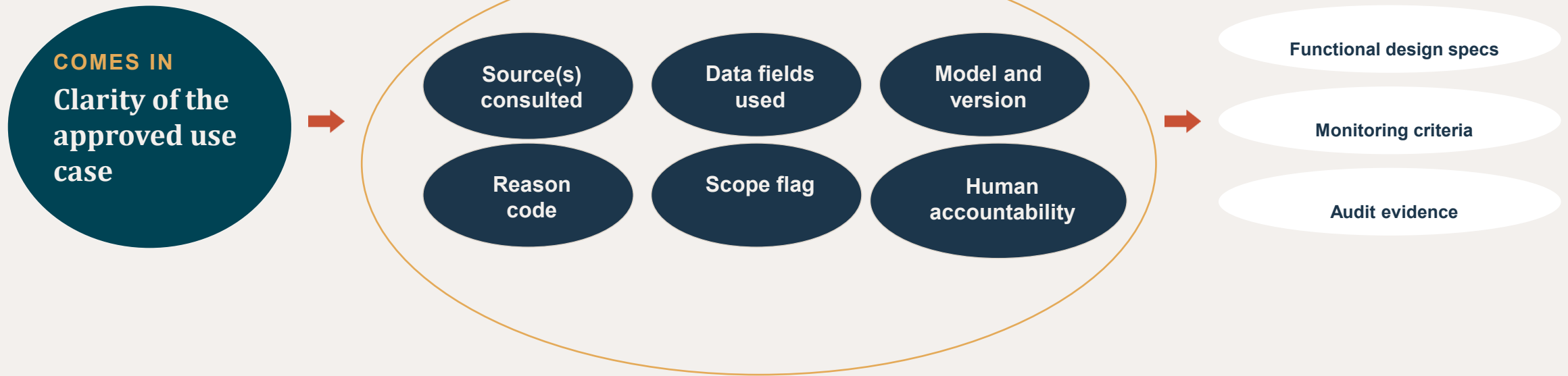
September - Introducing The Decision Record

Explainability, operationalized

		Documented Before: As Functional Requirements	Referenced Later: As Evidentiary Proof
1	Source(s) consulted	Specifies the authoritative source(s) from which the system can draw for its response.	Proof the decision is rooted in an approved source.
2	Data fields used	Specifies the allowed inputs for decisioning, excluding fields never appropriate for decision influencing.	Confirms only permitted inputs were used.
3	Model & version	Specifies the model type and version at the time of build (important for version control).	Ties a specific decision to a specific, reproducible state of the system.
4	Reason code	Specifies the list of reasons, and their approved definitions, the AI assigns to each of its decisions / outputs.	Ensures decision consistency and eliminates differing interpretations of reasons.
5	Scope flag	Defines operating boundaries for in-scope questions related to the use case.	Confirms the system only responded to queries it was built and approved to handle.
6	Human accountability	Identifies who owns decision review and issue escalation, and where on the loop spectrum they sit.	Proof of human intervention as designed.

The Decision Record is where requirements become proof

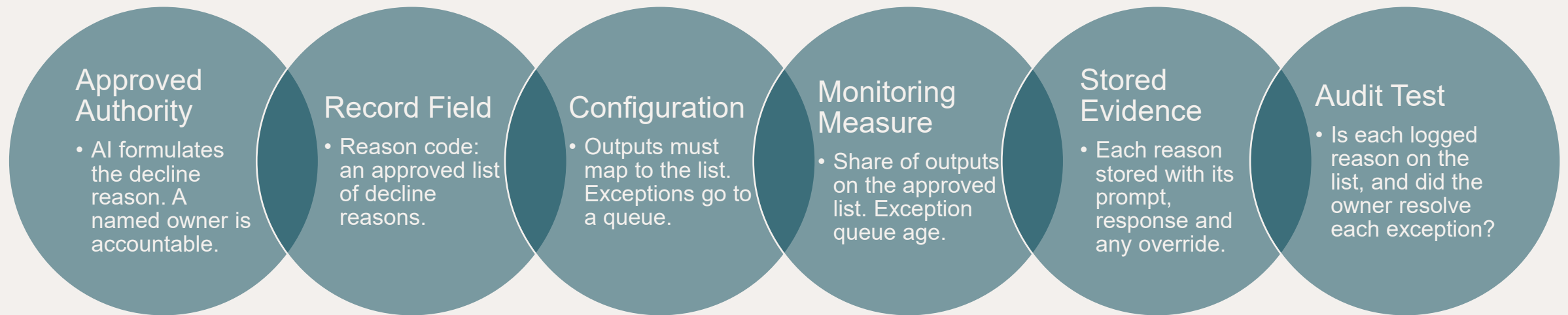
Decision Record: six fields



Written before build, it is **requirements**. Referenced after, it is **proof**. On paper alone, it assures nothing.

One decision, end to end

Illustration: a credit decline reason, followed through the chain.



What an AI audit actually tests

Design and configuration

- Question
 - Does what was built match the record?
- Audit Test
 - Compare the system's configuration to each field of record.

Execution

- Question
 - Do run-time logs show it was followed?
- Audit Test
 - Sample the logs against sources, reason codes and scope.

Accountability

- Question
 - Did a human review and act as designed?
- Audit Test
 - Trace exceptions and overrides to the named owner.

Then test the record itself: is it **complete, approved at the right level, and still current?**

Sampling the logs against the record is the AI equivalent of going to the warehouse.

Internal audit discipline, applied to AI

Familiar internal audit practice	Applied to AI
Risk-based planning	Rigor scaled to the use case's risk level
Defined criteria	The Decision Record
Evidence	Run-time logs and monitoring
Independence	Audit does not approve what it later audits

Audit is point-in-time, so monitoring runs beside it.

When you bought the system – the decisions it makes are still yours to prove

What you need from the vendor

- Model and version, with notice before either changes
- Run-time logs for your decisions, in a form you can compare to your record
- Retention for as long as the decision can be challenged
- Access for an independent party to test

Without it

- Your Decision Record goes out of date without anyone noticing
- Audit has nothing to sample
- The evidence is gone before anyone asks for it
- You are relying on the vendor's report about itself

If you can't get the evidence, **grant less authority**. The risk level should match **the proof** you can obtain.

Decision stewardship, revisited

The Decision Steward makes sure the point where organizational authority becomes AI authority is **deliberately configured, continuously justified and demonstrably trustworthy**.

Qualification

- Is this use case worth doing, and at what risk?
- Owns the approved use case

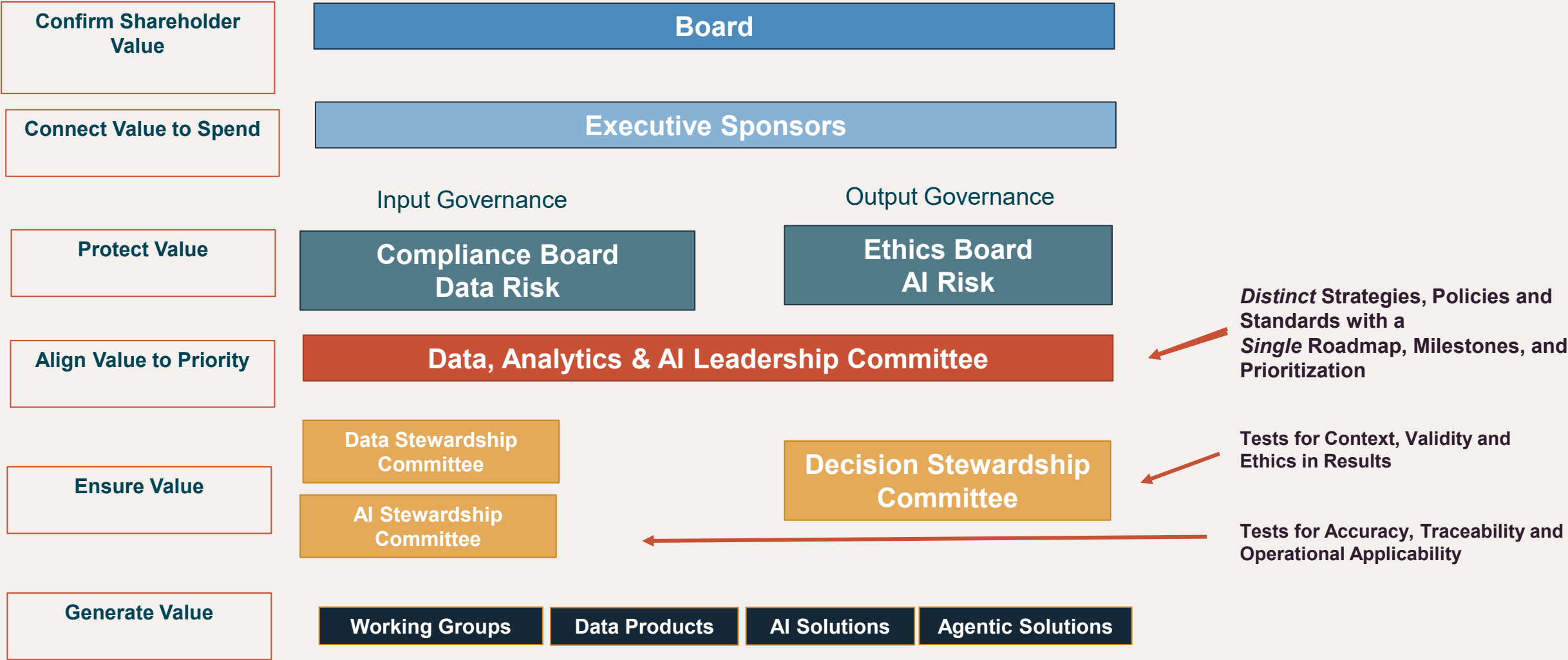
Delegation

- How much authority does AI get, for each decision?
- Turns the use case into the record's requirements

Operational

- Is the decision still holding in production?
- The steward produces the evidence. Internal audit tests it and reports to the board, not to the steward.

Decision-Making in an AI-Centric Organization



Which decisions are made where – applying the Data Decision Framework

Strategic direction	Board of Directors, Executive Sponsors	Set AI risk appetite. Receive assurance reporting.
Domain policies and requirements	Ethics Board, AI Risk, Compliance Board, Data Risk	Define risk levels and minimum assurance by risk level and limits on AI authority.
Departmental alignment	Data, Analytics and AI Leadership Committee	Approve qualification (go or no go). Prioritize and resource.
Operational decisions	Decision, AI and Data Stewardship Committees	Delegation and revalidation. Implement record specs. Data readiness.
Tactical daily work	AI solution pods, individuals	Monitoring, exception handling, evidence capture.
Outside the decision chain	Internal audit	Test conformance to the record. Never approve what it later audits.

Five questions to ask of any assurance claim

#	Question	Where we built it
1	Does it cover the organization, or a decision? Can you reconstruct one decision with it?	The audit tests
2	Does the evidence reconcile? Do the specification, the log and the human action agree for one named decision?	The audit tests
3	Who produced the evidence, and who checked it?	Decision stewardship and internal audit
4	Will the evidence still exist when someone asks?	When you bought the system
5	Can you get it from a vendor you don't control?	When you bought the system

Questions 1 and 2 ask whether anyone looked at the actual thing. Question 3 asks who checked the people who produced the evidence.

Where to start

One use case, one flow, not all of them.

1

- Pick one AI use case.

2

- Answer the qualification and delegation questions.

3

- Write the Decision Record.

4

- Configure at least one field, and agree what an independent party would test.



If an outside party asked you tomorrow to prove your highest-stakes AI decision, could you show that the use case was approved, what the system was specified to do, and that it did it?



Would they look at the warehouse, or at the paperwork?



Is your org **AI-Ready?**
Take our 3-minute
assessment, get instant
results



Kelle O'Neal | Founder & CEO
Lisa Wintrick | Executive Advisor

