FIRST SAN FRANCISCO PARTNERS

THE RIGHT TO EXPLANATION

Meeting Regulatory Demands for Interpretable AI

AIGOV | LEADING AI Governance

DATAVERSITY®

Welcome to Leading AI Governance – Session 9

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

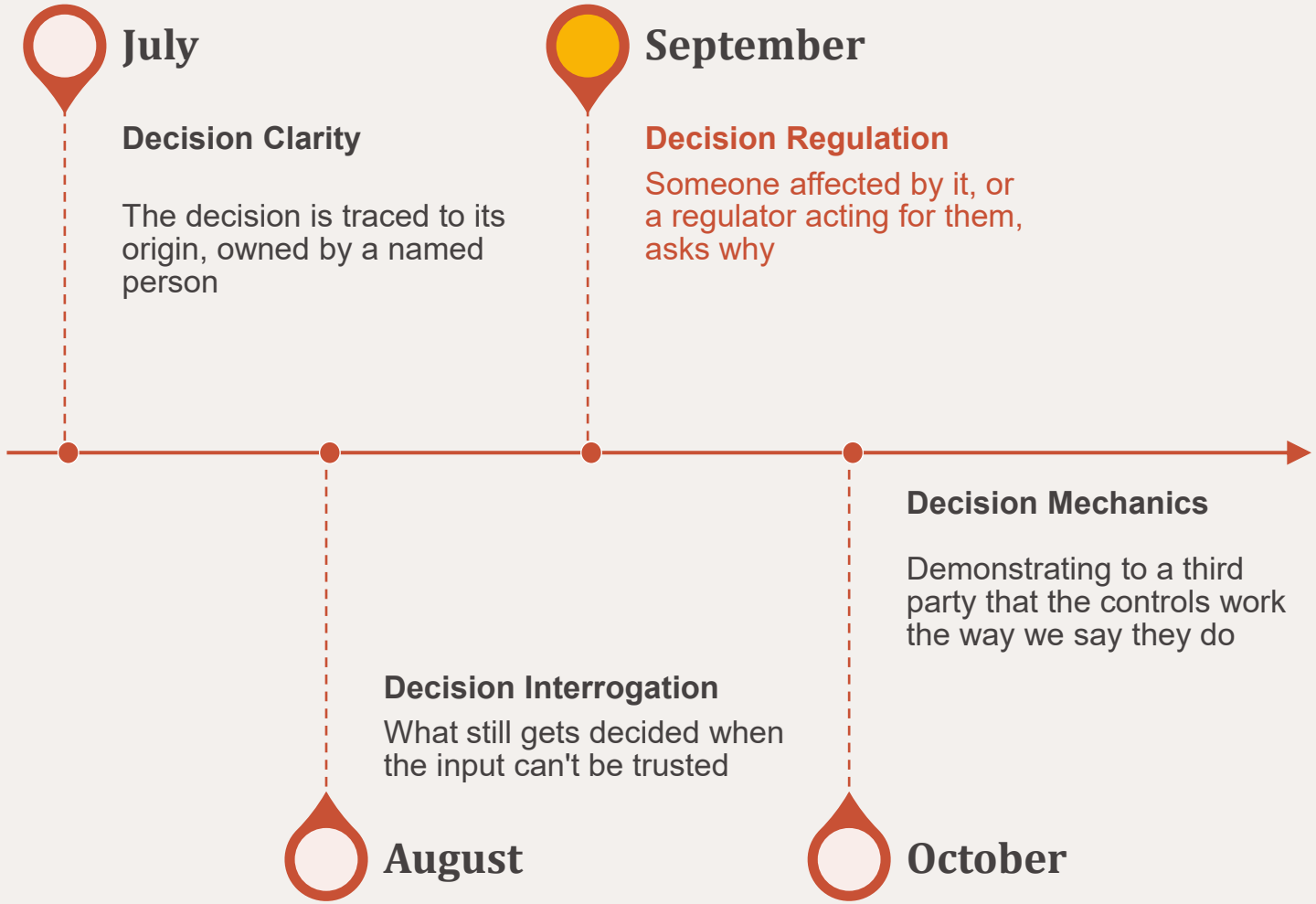
Founder and CEO,
First San Francisco Partners

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

Where This Session Fits

This session covers what we owe, to whom, and in what form.



AIGOV | LEADING AI Governance

Two different questions, two different audiences

Why this matters before we define anything

A regulator asks why we declined this applicant.

Can we answer for this case, and does the answer hold up?

The Board asks if we can answer that question every time.

Can we say yes with confidence?

Today's Topic

PROVIDING INTERPRETABILITY AND EXPLAINABILITY OF AI SOLUTIONS TO KEY STAKEHOLDER GROUPS

What You Will Leave With

How to answer the regulator's question

What "explainable" actually requires, and how to tell if you already have it

How to answer the Board's question

The difference between answering once and being able to answer every time, and what makes that repeatable

Where the gap between legal responsibility and technical reality actually sits

And why closing it is a governance decision, not a modeling one

Thesis

Real technical limits exist that impede model *explainability* and the *interpretability* of that account.

To counterbalance this, anchor on something governed, stable, and meaningful outside of the model.

THE PROBLEM

A model's account of its reasoning is a plausible reconstruction – not a real-time record of its logic. This immediately limits interpretability.

THE SOLUTION

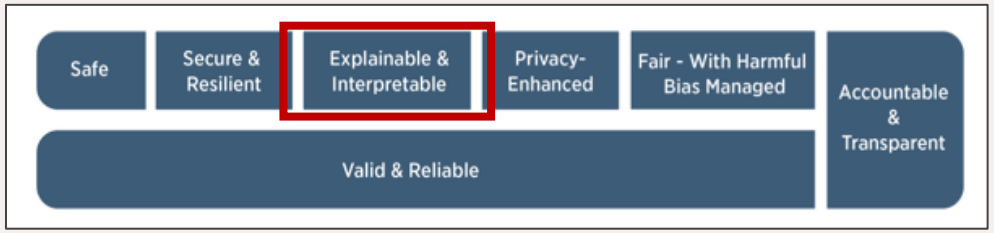
A **Decision Record** that serves as both a design input to the AI system and an answer key for questions from each stakeholder group.

THE LIMITATION

The Decision Record manages – rather than eliminates – AI risk. Organizations must therefore exercise greater diligence over controllable variables surrounding the AI model.

Distinguishing explainability vs. interpretability

Sequence, not synonyms



Source: NIST AI 100-1 Risk Management Framework

EXPLAINABLE

Answers *how* a decision was made in the system.

The model's account of its own mechanism.

MODEL CAPABILITY

INTERPRETABLE

Answers *why* a decision was made by the system.

How a person makes sense of the model's account.

USER CAPABILITY

An example

EXPLAINABILITY

How was this decision made?

The model flagged debt-to-income ratio at 30%, below the 35% approval threshold.

INTERPRETABILITY

Why does that matter, and what does it mean to me?

That ratio, *not* credit history, is why you were approved for the loan. Lowering monthly debt or documenting more income assists in improving the debt-to-income ratio.

Explainability is four requirements, not one

PRINCIPLE 1

Explanation

Evidence or reason(s) for outputs or processes.

NECESSARY, NOT SUFFICIENT

PRINCIPLE 2

Meaningful

Understandable to the intended consumers.

REQUIRES DELIBERATE DESIGN

PRINCIPLE 3

Knowledge Limits

The system only operates under conditions for it was designed.

REQUIRES DELIBERATE DESIGN

PRINCIPLE 4

Explanation Accuracy

The explanation correctly reflects the reason or process for generating the output.

Matches what the model computed

UNPROVABLE TODAY

Consistent, approved, and defensible

PROVABLE NOW

*Three are governance problems, solvable without inspecting inside the model. The fourth splits into two – **but one of them is attainable.***

The next question on the table for the courts

Mobley v. Workday (May session)

A vendor's model doesn't shield the employer. You can't point at the AI and say it wasn't you.

Navy Federal Mortgage Discrimination Litigation, 4th Cir., Feb. 9, 2026

Plaintiffs allege Navy Federal's underwriting algorithm decides creditworthiness using variables and weights the company keeps secret, even from applicants it denies. Claims allowed to proceed; nothing about liability decided.

Navy Federal is being asked to produce *exactly what the fourth requirement called attainable* - not the model's internals, but a consistent, defensible account of what drove the decision.

Revisiting July's session - a credit decline

A bank uses AI to draft adverse-action letters that tell a customer why they were declined

GROUNDED

AI can only cite decline reasons from an approved list tied to the decision data, not free text it composes.

TRACED

Every letter records the model version, the data fields used, and the template.

OWNED

A named owner approves the process and owns the exception queue. *A letter that can't map to an approved reason doesn't go out. It routes to a person.*

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."

EXPLAINABILITY, OPERATIONALIZED

The Decision Record

		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 a key artifact of Decision Stewardship that defines AI system boundaries.

Since we can't govern the model, we govern everything around it.

It provides the operating conditions for *explainability*, and the information, in consistent language, for stakeholder *interpretability*.

DESIGN & CONFIGURATION

What the organization decided and built

EXECUTION

What the system logged as having occurred at run-time

ACCOUNTABILITY

What human-based decision was made on performance and output

REAL-WORLD APPLICATION

Where to start

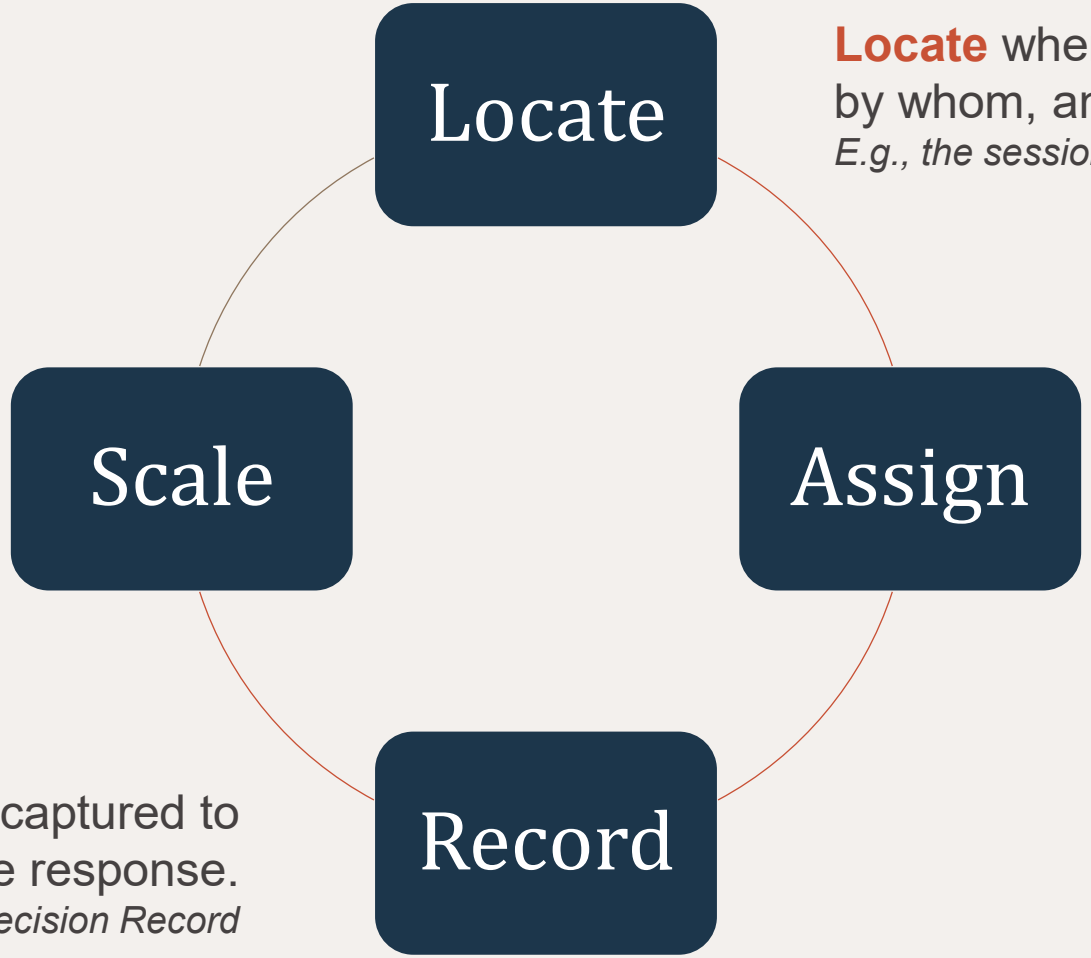
This is the same test as the competency question, just run backward. Earlier we asked what questions a decision flow must answer before you build it. This is that question but asked of something you've already built.

Same discipline, applied retroactively.

2 OF 4 STEPS COMPLETE

- [x] **Pick one decision flow, not all of them.**
- [x] **Write down who asks about it, and what they actually ask.**
- [ ] Try to write the answers today, before you (re)build anything.
- [ ] Any resulting gap is an opportunity: fixable with a decision record.

Zooming out – a consistent framework across sessions



Locate where the decision is made, by whom, and how.
E.g., the session opening questions

Assign an owner for the decision and consulted parties.
E.g., the business process / department owner

Record what needs to be captured to guide, scope, and direct the response.
E.g., the competency questions > Decision Record

Scale by governing the decision flow, not individual decisions.

AIGOV | LEADING AI Governance

Session recap

Technical limitations exist on explainability

- Models provide plausible reconstruction only
- Full explainability cannot be achieved
- This impedes interpretability

TECHNOLOGY

The Decision Record governs around the model

- Provides a defensible basis for model behavior
- Aids in consistency of interpretability
- Also provides scoping and configuration requirements

POLICY

Competency questions are the input to the Decision Record

- Answers questions of the system and to the system
- Input to the decision record
- Also serves as valuable scoping tool for AI builds
- No approved answer(s) means the decision flow isn't explainable yet

PROCESS

Decision Stewardship produces the Decision Record

- The Decision Record is a key artifact of Decision Stewardship
- Ensures the use case is valid for AI decisioning
- Facilitates the completion of the Decision Record
- Maintains it over time

PEOPLE

AIGOV | LEADING AI Governance

Back to the Boardroom

What we can promise, and what we can't

A regulator asks why we declined this applicant.
Can we answer for this case, and does the answer hold up?

The Board asks if we can answer that question every time.
Can we say yes with confidence?

We can't tell you exactly how the model reasoned. We can tell you the basis for decisioning, the operating environment, and human involvement - **every time**, not just this once.

AIGOV | LEADING AI Governance

FIRST SAN FRANCISCO PARTNERS

Let's Stay Connected
Scan the QR Code

Lisa Wintrick | Executive Advisor
Stephanie Paradis | Principal Consultant

Session 10 — Building a Framework for AI Assurance | October 6, 2026

AIGOV | LEADING AI Governance