

— AUDIT · RISK · GOVERNANCE · AI

AI risk isn't only model risk — it's **Reconstructability Risk**.

How to make AI-assisted decisions reconstructable, auditable,
and defensible.

CEO, Invariant Engineering Group · CEO, LexLatam.ai

JD, MBA · 20+ years software engineering · MIT Sloan AI for Business · ISACA Speaker

Most AI failures are not only model failures.

They are failures of —

↳ **Workflow design**

📁 **Evidence preservation**

👤 **Human review**

🕒 **Approval discipline**

📊 **Decision lineage**

🔍 **Auditability**

THE QUESTION THIS TALK IS BUILT AROUND

Can we reconstruct how an AI-assisted decision was made **six months later?**

Three workflow breakdowns → one core argument.

1 Three public workflow breakdowns

2 The Auditability Gap

3 The Six-Month Reconstruction Test

4 Audit Grill-Me

5 Audit-Driven Development

6 AI Workflow Auditability as an emerging capability

7 Practitioner pressure-testing

~40 min lecture · 15 min Q&A · 5 min feedback

What this talk is — and is not.

THIS IS NOT

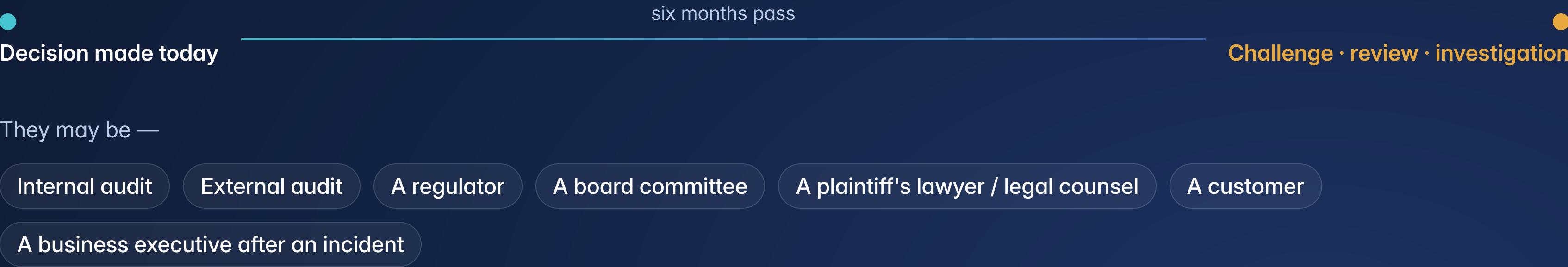
- × An anti-AI talk
- × An anti-model-risk talk
- × Legal advice
- × A product pitch
- × A replacement for existing AI governance frameworks
- × An announcement of a new certification registry

THIS IS



A practical audit conversation about whether AI-assisted workflows can be **evidenced and reconstructed.**

Six months later, someone asks: "Why did this AI-assisted decision happen?"



The auditor's question is different.



Governance asks

Should we allow this?



Engineering asks

Can we build this?



Risk asks

What could go wrong?



Compliance asks

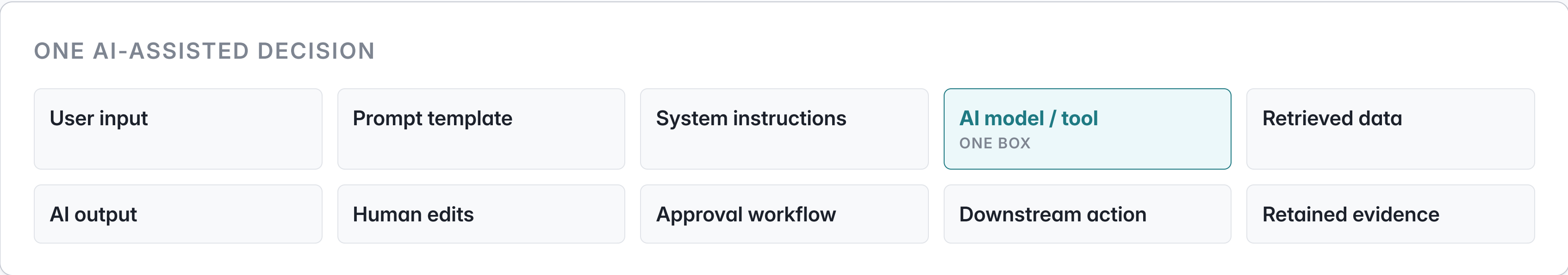
What obligations apply?



Audit asks

Can we prove what happened?

The model is one component — not the whole control environment.



Model risk covers one box. **Workflow risk** covers the other nine — and the seams between them.

Mata v. Avianca — fabricated authorities in an official filing.



Attorneys filed non-existent judicial opinions generated by ChatGPT.



The court: nothing inherently improper about using a reliable AI tool — but attorneys retain a **gatekeeping responsibility**.



The failure compounded — they **stood by the fake citations** even after the court questioned them.

 **Workflow lesson** — what mechanism should have prevented unverified AI output from becoming an official filing?

Moffatt v. Air Canada — liable for the chatbot's answer.



The BC Civil Resolution Tribunal held the airline **liable for inaccurate chatbot information** about bereavement fares.



A specific tribunal decision, not universal precedent — but the control implications are practical.

THE CONTROL QUESTIONS



Who approved the data feeding the chatbot?



How was it tested?

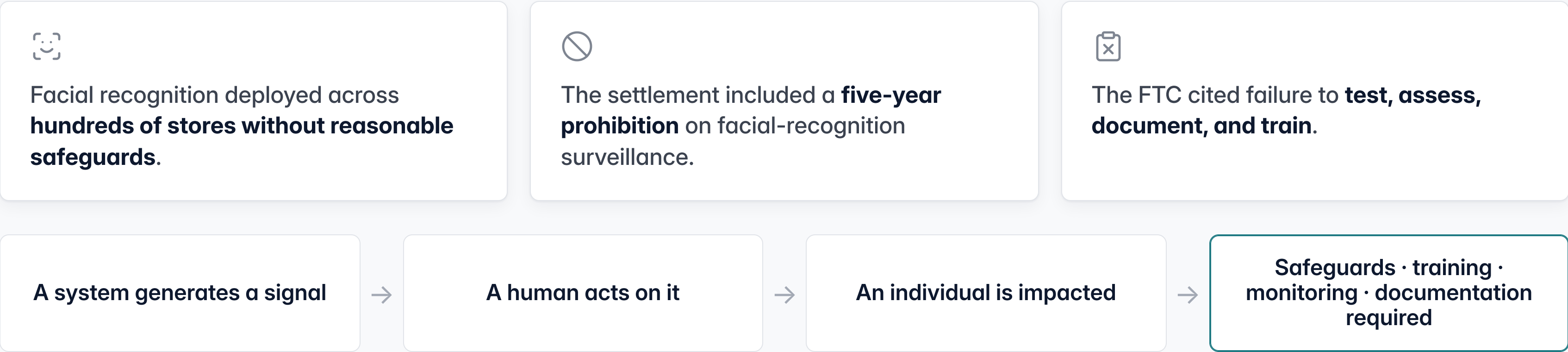


Was it aligned with official policy?



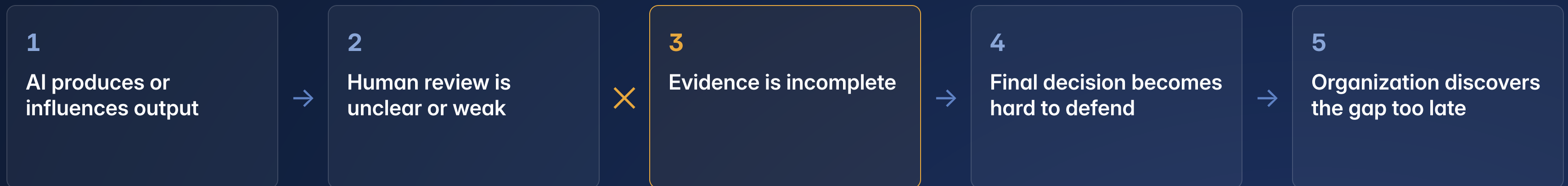
Were outputs logged and monitored?

FTC / Rite Aid — surveillance without safeguards.



The pattern: AI output becomes organizational action before controls and evidence are mature.

The break is usually around the evidence.



The failure is often not the AI output. It is the human–AI workflow around the output.

Model risk lives inside workflow risk.

WORKFLOW RISK ASKS

MODEL RISK ASKS

- Is the model accurate?
- Is it biased?
- Does it hallucinate?
- Was it tested?
- Is it explainable?

- Who used it?
- What data was used?
- What did the human do?
- Can we prove it?
- What did they ask?
- What did the AI produce?
- Who approved the action?

WORKFLOW RISK

The risk that an AI-assisted process cannot be **controlled, evidenced, reconstructed, challenged, or defended** — from initial input to final decision.

Sharpened further: **reconstructability risk** — the risk that the decision path cannot be proven after the fact.

| If you cannot reconstruct the decision, you may not be able to defend the decision.

The gap between paper and proof.



Governance says what **should** happen.
Auditability proves what **did** happen.

Every governance artifact has an auditability question.

GOVERNANCE ARTIFACT	AUDITABILITY QUESTION
AI policy	Did the workflow follow it?
Model inventory	Which system was actually used?
Human-in-the-loop rule	What did the human actually review?
Vendor contract	Can we access the logs?
Approval matrix	Was approval actually obtained?
Retention policy	Was evidence actually retained?

 POLL 1

Could your organization reconstruct a material AI-assisted decision from six months ago?

- ☐ Yes, with complete evidence
- ☐ Partially, with manual effort
- ☐ Probably not
- ☐ We do not know
- ☐ We are not using AI in material workflows yet

If the answer is "**we do not know**," who would know?

The Six-Month Reconstruction Test

Select **one material AI-assisted decision** from six months ago. Can you reconstruct it — end to end?

- ✓ Who initiated it?
- ✓ What prompt or input was submitted?
- ✓ What output was produced?
- ✓ Who approved it?
- ✓ Where does the evidence live?
- ✓ What AI system was used?
- ✓ What data or documents were used?
- ✓ What human review occurred?
- ✓ What final action was taken?

If you cannot reconstruct the path, it does not mean the decision was wrong. **It means the decision is indefensible.**

Why six months?

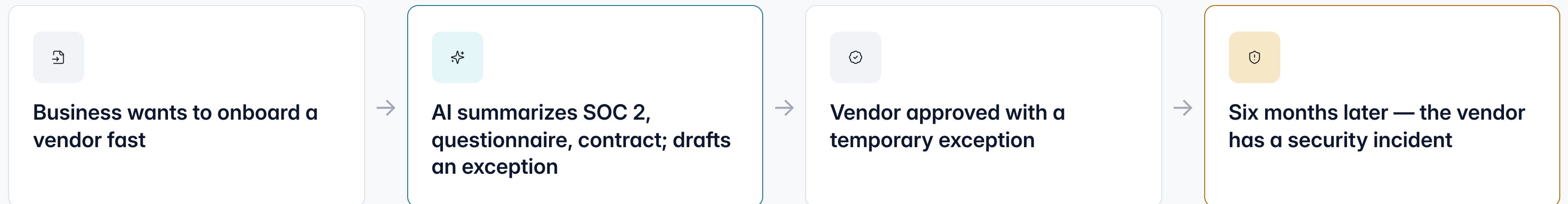
Long enough for —

- 🕒 **Memory to fade**
- ✍️ **Prompts to change**
- 🗄️ **Vendor logs to expire**
- 👤 **Employees to change roles**
- 🔄 **Models to update**
- 👉 **Informal evidence to disappear**



**Not a universal legal retention rule — a practical
audit-readiness stress test.**

AI-assisted vendor risk exception.



🔍 Audit asks: **Why was the exception approved?**

What audit needs to know.

01

Who requested the exception?

02

What AI tool summarized the documents?

03

Which documents were used?

04

What did the AI recommend?

05

Did the reviewer verify the summary?

06

What risks were accepted?

07

Who approved the exception?

08

Were remediation conditions tracked?

09

Was the exception still open at the time of the incident?

 AI Decision Evidence Packet

For material AI-assisted decisions, preserve enough evidence to reconstruct —

- | | | | |
|----------|--|-----------|--------------------------------------|
| 1 | Business context | 6 | AI output |
| 2 | User identity | 7 | Human review |
| 3 | Prompt or input | 8 | Approval chain |
| 4 | System instructions or template | 9 | Final decision |
| 5 | Data sources | 10 | Retention and access controls |

What each evidence element proves.

EVIDENCE ELEMENT	WHAT IT PROVES
Business context	What decision was being made
Prompt / input	What the human asked
Data sources	What the AI relied on
AI output	What the system produced
Human review	What judgment was applied
Approval chain	Who authorized the action
Final decision	What actually happened
Retention controls	Whether evidence can be trusted later

Do not log everything forever.

Poor evidence preservation can create —

Privacy risk

Privilege risk

Confidentiality risk

Discovery risk

Cybersecurity risk

Retention risk

Unmanaged sensitive-data repositories



THE BETTER QUESTION

What evidence is **necessary, proportionate, protected,** and retained for the right period?

Where reconstruction usually breaks.

⚠ Prompt not retained

⚠ Output copied without context

⚠ Source documents not linked

⚠ Model or tool version unknown

⚠ Human edits not tracked

⚠ Approval happened outside the system

⚠ Vendor logs unavailable

⚠ Retention period too short

⚠ Evidence retained without access controls

Audit Grill-Me — a 90-second reconstruction drill.

Ask the workflow owner —

Show me the prompt.

Show me the output.

Show me the source data.

Show me the human review.

Show me what changed.

Show me the approval.

Show me the model / tool version.

Show me the final decision.

Show me the retention rule.

Human-in-the-loop is not a control by itself.

A real human review control needs evidence of —

- ✓ What the human reviewed
- ✓ What they changed
- ✓ What they escalated
- ✓ Whether they had authority
- ✓ Did the interface encourage independent evaluation — or automation bias and rubber-stamping?
- ✓ What standard they applied
- ✓ What they rejected
- ✓ What they approved
- ✓ Whether they had competence



Human-in-the-loop is not enough if the human is only **near the loop.**

 POLL 2

Where is reconstruction most likely to fail in your organization?

☐ User prompts

☐ Retrieved information / knowledge sources

☐ Human review

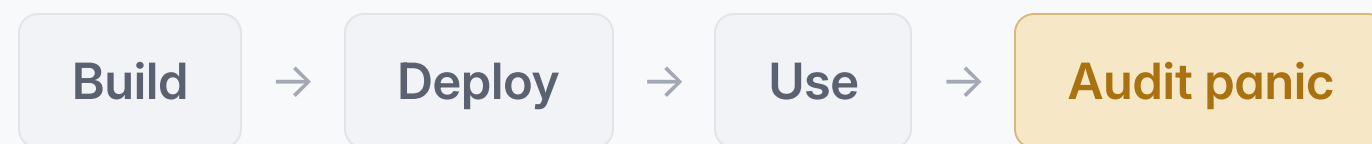
☐ Approval and sign-off records

☐ Logging and audit trail

☐ I don't know

Auditability should not be retrofitted.

OLD PATTERN



BETTER PATTERN



Auditability should not be retrofitted after the workflow is already operating.

AUDIT-DRIVEN DEVELOPMENT

The practice of engineering AI-assisted workflows so that material decisions are **inherently reconstructable from day one**.

- ✓ Define evidence requirements before deployment
- ✓ Test reconstruction before launch
- ✓ Build controls into the workflow
- ✓ Treat auditability as part of "done"

A new pillar beside the "by design" disciplines.

Test-driven development

Can we prove the system behaves as expected?

Security by design

Can we reduce attack surfaces early?

Privacy by design

Can we minimize and protect personal data?

Audit-Driven Development

Can we reconstruct the decision later?

Auditability embedded across the lifecycle.



Sample control objectives.



Decision lineage



Prompt provenance



Output preservation



Source traceability



Human review evidence



Approval-chain integrity



**Model / tool version
traceability**



Exception documentation



Retention & access control



Reconstruction testing

Existing frameworks feed into workflow evidence.

NIST AI RMF

Govern · Map · Measure · Manage

ISO/IEC 42001

AI management system

COBIT

Governance & management objectives

COSO

Internal control over generative AI

OWASP LLM Top 10

Technical risk patterns for LLMs

ISACA AAIA

AI audit professionalization



Workflow evidence

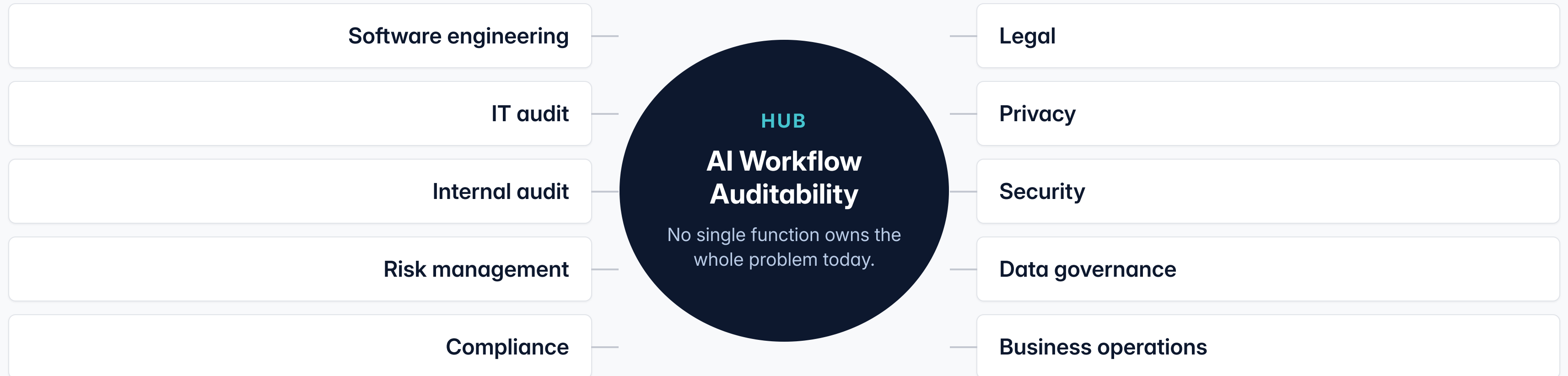
Frameworks set governance expectations. AI Workflow Auditability asks whether the workflow can produce the evidence.

AI WORKFLOW AUDITABILITY

The practice of designing, assessing, and improving AI-assisted workflows so material decisions can be **reconstructed, evidenced, challenged, and defended.**

| Core question: **Can we prove how the decision happened?**

Where this specialty sits.



What this specialty would do.



Define evidence requirements with engineering during design



Establish decision lineage for material workflows



Verify vendor logging capabilities during procurement



Balance retention with privacy and privilege risk — with legal and privacy teams



Ensure the business can prove its automated processes before stakeholders demand it

Emerging specialty or required competency?

🔗 A standalone professional specialty

🔗 A sub-discipline of IT audit

🎓 A required AI governance competency

🔗 A software engineering practice

🔗 A GRC control domain

👥 A cross-functional assurance role

The name matters less than **the capability**.

What a credible movement would need.



Shared glossary



Evidence Packet template



Reconstruction Test worksheet



Audit Grill-Me question bank



Control objective library



Sample audit program



Maturity model



Vendor questionnaire



Framework crosswalk



Practitioner working group

Help pressure-test the idea.

- ② What evidence would be sufficient?
- ② Who should own this?
- ② Where would your organization fail first?
- ② What evidence would be excessive?
- ② How should auditors test it?
- ② What should this capability be called?

NOT A SALES PITCH

A call to define a practical control discipline for reconstructable AI-assisted workflows.

Q&A

What would you challenge?

→ Is this actually new?

→ Where does the concept break?

→ Who should own it?

→ What evidence is enough?

→ What evidence creates new risk?

→ What would fail first in your environment?

Share your feedback.

- 1 Does your organization pass the Six-Month Reconstruction Test?
- 2 Where would reconstruction fail first?
- 3 Does "AI Workflow Auditability" describe a real governance gap?



Scan to respond

AI governance tells us what **should** happen.
AI Workflow Auditability helps us prove what **actually** happened.

THE NEXT PRACTICAL FRONTIER

Reconstructable human–AI workflows.

Audit-Driven Development is how we build them.



Examples used as illustrations.

Mata v. Avianca	Court sanctions order involving non-existent judicial opinions generated through ChatGPT and submitted in a legal filing.
Moffatt v. Air Canada	BC Civil Resolution Tribunal matter involving inaccurate chatbot information about bereavement fares, as summarized by the ABA.
FTC Rite Aid settlement	FTC alleged Rite Aid deployed facial recognition without reasonable safeguards and failed to test, assess, document, or train adequately.
SEC AI-washing actions	SEC announced settled charges against two investment advisers for false and misleading statements about purported AI use.

 Use these as workflow / control illustrations, not broad legal conclusions.

— REGULATORY & FRAMEWORK SIGNALS

Relevant signals.

NIST AI RMF Core

Govern, Map, Measure, Manage.

ISO/IEC 42001

Requirements for an AI management system.

EU AI Act

High-risk AI logging, record-keeping, and human oversight concepts.

ISACA AAIA

AI governance and risk, AI operations, and AI auditing tools and techniques.

COSO GenAI guidance

Internal-control framing for generative AI.

OWASP LLM Top 10

LLM application risk patterns.

Packet template.

SECTION	REQUIRED EVIDENCE
Decision context	Business purpose, workflow name, risk rating
Actor identity	User, reviewer, approver
AI interaction	Prompt / input, template, system instructions
Data basis	Source documents, databases, retrieval results
System traceability	Tool, model, version, configuration, vendor
Output	Raw AI output and final edited version
Human review	Reviewer action, changes, rationale, escalation
Approval	Approval chain, timestamp, exception record
Final action	System transaction, decision record, communication
Retention	Storage location, retention rule, access controls

SAMPLE AUDIT PROCEDURES

For a sample of material AI-assisted decisions.

- 1 Obtain the final decision record.
- 2 Identify the AI-assisted workflow used.
- 3 Trace the decision back to prompt / input.
- 4 Verify source documents or data used.
- 5 Obtain AI output and compare to final decision.
- 6 Identify human reviewer and reviewer action.
- 7 Verify approval chain and authority.
- 8 Confirm model / tool / version traceability.
- 9 Review exceptions, overrides, and escalations.
- 10 Confirm evidence retention and access controls.

— VENDOR DUE-DILIGENCE QUESTIONS

AI vendor auditability questions.

- ② What logs are available?
- ② How long are logs retained?
- ② Can logs be exported?
- ② Can prompts and outputs be linked to users and timestamps?
- ② Can logs support audit, investigation, or regulatory inquiry?
- ② Can model or tool versions be identified after the fact?
- ② What happens when the model changes?
- ② Can source documents or retrieval results be traced?
- ② Are admin actions logged?
- ② What privacy and security protections apply to retained evidence?

— AI WORKFLOW AUDITABILITY MATURITY MODEL

From unaware to optimized.

0 Unaware

AI use is informal; no reconstruction expectation.

1 Ad hoc

Some evidence exists, but inconsistently.

2 Documented

Policies exist; selected workflows retain artifacts.

3 Controlled

Material workflows have evidence requirements and owners.

4 Tested

Reconstruction is periodically tested through sampling.

5 Optimized

Auditability is embedded into SDLC, vendor review, monitoring, and continuous improvement.

— NAMING POLL

Which term best describes the missing capability?

☐ AI Workflow Auditability☐ Human–AI Decision Assurance☐ Decision Lineage Assurance☐ Audit-Ready AI Workflows☐ Reconstructable AI Workflows☐ AI Evidence Engineering☐ Something else

What term would make business, audit, legal, and engineering teams all understand the same problem?