Internal Audit AI Testing Framework

1. Purpose of This Framework

Artificial Intelligence introduces **new types of risk**, **new control expectations**, and **new assurance challenges** that traditional audit approaches do not fully address.

This matrix helps internal auditors:

- Understand where AI risk appears across the business
- Articulate control objectives for AI systems
- Identify key failure modes and red flags
- Align assurance work with recognized AI governance frameworks
- Determine testing focus based on risk/materiality

2. Where AI Shows Up in Internal Audit Work

Internal audit may encounter AI in four ways:

A. AI embedded in processes

- Forecasting tools
- Fraud detection
- Routing/triage
- Scoring models

B. AI augmenting audit procedures

- Generating test scripts
- Summarizing workpapers
- Extracting text

C. AI embedded within enterprise systems

- ERP system AI
- HRIS AI features
- Procurement tools
- Cyber detection engines

D. Custom AI models built by data science teams

- Predictive risk models
- Chatbots
- Natural Language Processing (NLP) analysis engines
- Computer vision models

This framework applies to all four categories.

3. Core AI Assurance Domains

Below are **the eight domains** every internal audit should consider when AI is in scope.

	Domain	*Risk Examples*	*Control Objective*
1	**Data Quality & Data Governance** AI outputs depend entirely on data inputs.	• Incomplete or biased data • Poor lineage documentation • Uncontrolled data transformations	Ensure data feeding AI is complete, accurate, authorized, and fit for purpose.
2	**Model Design & Development** How the model is built.	• Incorrect algorithms • Overfitting • Misalignment with business purpose	Models are designed using sound methodologies aligned to documented requirements.
3	**Model Validation** Independent confirmation the model works as intended.	• Underperforming models • Unvalidated assumptions • Unexplained behavior	Independent validation confirms acceptable accuracy, robustness, fairness, and reliability.
4	**Explainability & Transparency** Stakeholders must understand how the model works.	• "Black box" logic • Inability to justify outcomes • Confusing or misleading explanations	AI output and decision logic must be sufficiently explainable to users.
5	**Drift Monitoring & Ongoing Model Performance** Models degrade over time.	• Model drift • Data drift • Over-time inaccuracies	Model performance is continuously monitored and retrained as needed.
6	**Access, Security & Change Management** AI tools must be secured.	• Unauthorized changes • Model poisoning • Shadow AI usage	AI models and tools are protected with role-based access and formal change control.
7	**Responsible Use & Governance** AI must be used ethically, safely, and legally.	• Bias • Hallucinated outputs • Misuse of confidential data • Policy violations	Policies govern acceptable AI use, dataset handling, testing, deployment, and oversight.
8	**Output Quality & Human Oversight** AI is fallible. Humans must review outputs (HITL).	• Inaccurate outputs • Unchallenged hallucinations • Overreliance on results	AI outputs undergo human review proportionate to risk and materiality.

4. The Internal Audit AI Testing Matrix (Conceptual View)

Below are the **summarized, high-level risk/control/testing lens** auditors should use.

Domain	***Key Risks***	***Control Objective***	***Testing Focus***	***Red Flags***
Data Quality	Incomplete, biased, corrupted data	Data is accurate, complete, authorized	Trace data lineage, review data controls	No lineage, unclear sources
Model Development	Incorrect logic, weak design	Sound, documented design	Review requirements, methodology	"Black box" model
Model Validation	Poor performance, unvalidated models	Independent validation	Review validation reports	No validation or outdated
Explainability	Inability to justify outputs	Transparent logic	Assess interpretability	"We can't explain why it did that"
Drift Monitoring	Degraded accuracy	Performance continuously monitored	Review drift logs	No drift tracking
Security & Change Control	Unauthorized access, shadow AI	Controlled access, documented changes	Review logs and approvals	No access logs
Responsible Use	Data leakage, policy violations	Compliant, ethical AI use	Check policy alignment	Prompts contain confidential data
Output Quality	Incorrect results	Human review ensures accuracy	Reperform outputs	Staff using AI without validation

This table provides a **quick lens** for scoping and planning audits.

5. Guidance on Scoping AI Audits

Use this decision tree:

Question 1: Does AI directly affect financial reporting, compliance, customer decisions, or high-risk processes?
→ *If YES, full AI assurance required.*

Question 2: Does the system rely heavily on automated scoring, classification, or predictions?
→ *If YES, include model risk domains.*

Question 3: Is the AI used only for productivity support (drafting, summarization)?
→ *If YES, focus on data governance, prompt governance, and human review.*

Question 4: Is the AI vendor-provided?
→ *If YES, obtain vendor SOC, ISO, and AI transparency reports.*

Question 5: Is this AI experimental/pilot/staff-driven?
→ *If YES, scrutinize usage, data leakage, and oversight.*

6. Framework Alignment (Conceptual Mapping)

Refer to Handout 5 "Responsible AI Governance for IA" for more detail

NIST AI RMF

- Maps directly to Data, Governance, and Technical Controls
- Good for risk taxonomy

EU AI Act

- Classification of risk levels
- Human oversight expectations

ISO/IEC 42001 (AI Management System)

- Governance, documentation, lifecycle expectations

COSO + IIA Standards

- Control objectives and risk assessment alignment

7. Red Flags That Should Trigger Deeper Testing

- "We can't explain how the model makes decisions."
- No documentation of training data.
- No drift monitoring.
- No independent validation.
- AI outputs bypass human review.
- Prompts contain sensitive information.
- End-users treat AI outputs as *facts*.
- Shadow/unauthorized AI use in business units.

Refer to Handout 4 "*AI Audit Testing Toolkit* which provides practical, operational test scripts and templates.