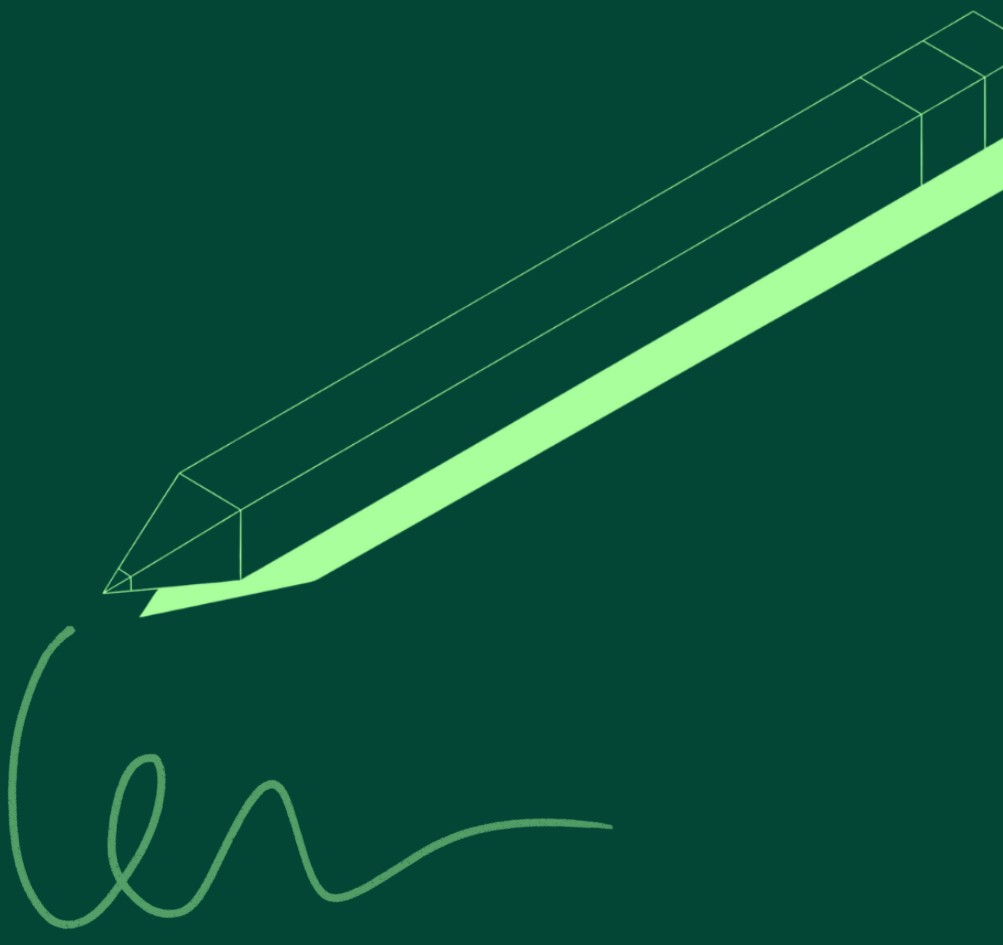
TRULLION

eBook

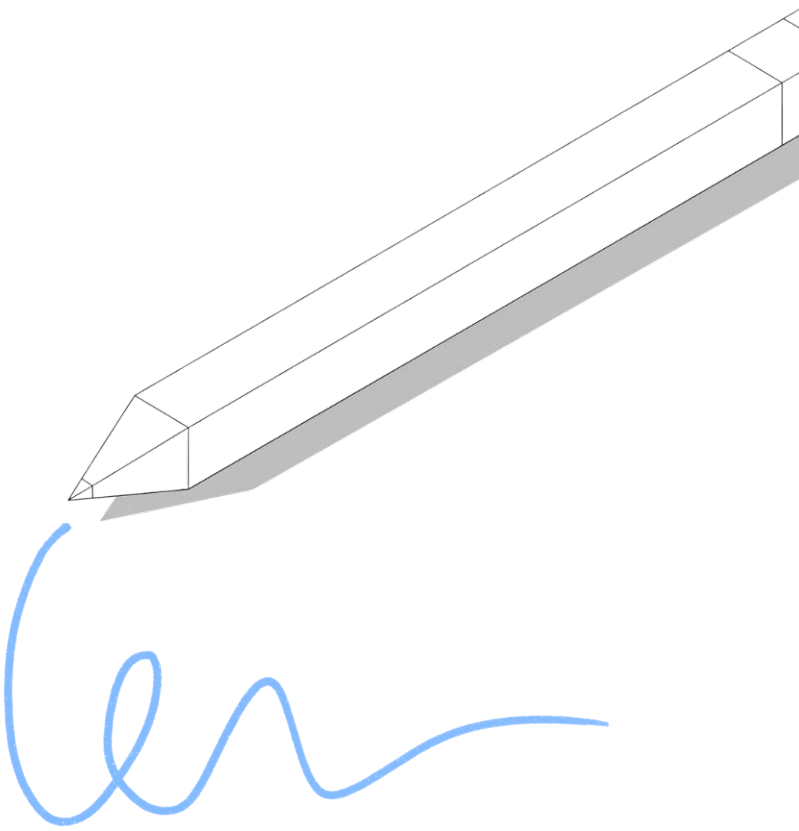
From Pencils to Agents: The Audit AI Readiness Assessment

Benchmark Your Firm Across Six Levels of Testing Automation



Introduction

Get a realistic look at where your firm's AI readiness stands today. Learn how to evolve your usage from foundational automation to autonomous agentic testing.



Executive Summary: AI in Audit Is Accelerating, But Scaling Is Rare

AI adoption is rising across industries, but scalable impact remains limited.

- Nearly **9 in 10 organizations** report AI use in at least one business function.
- Yet only **23% are scaling agentic AI**, while many remain in experimentation mode. Source: McKinsey State of AI Report

More importantly: most AI initiatives never reach production.

- Only **10–15% of AI pilots scale into long-term systems.** Source: Forrester commentary via Economic Times
- In 2025, **42% of organizations abandoned most AI initiatives**, often due to unclear ROI and governance gaps. Source: WorkOS Enterprise AI Failure Patterns

For audit firms, this gap matters more.

Audit AI must preserve:

- Evidence integrity
- Reviewer oversight
- Documentation quality
- Regulatory defensibility

Regulators have emphasized the need for documentation showing what was done, by whom, and based on what evidence (PCAOB AS 1215). At the same time, reporting highlights that many large firms are investing heavily in AI but have not formalized how they measure its impact on audit quality. Source: Financial Times

When implemented correctly, AI can enable:

- Full-population testing
- Automated matching and recalculation
- Faster anomaly detection
- Stronger risk identification

The difference between stalled pilots and scalable transformation is readiness.

AI in Audit: Not “If” but “How”

AI is already reshaping audit workflows — from data ingestion to substantive testing to documentation.

But successful adoption does not come from skipping to autonomous agents.

It comes from building:

- Connected workflows
- Structured, reusable templates and workpapers
- Traceable documentation
- Governed oversight

Enterprise research consistently shows that AI fails when workflows and data foundations are weak. Nearly half of enterprise AI projects underperform due to poor data readiness and integration.
Source: Fivetran Data Readiness Report

Audit teams that modernize intentionally — rather than experiment broadly — are the ones positioned to scale.

The Core Idea: Audit Testing Evolves in Levels

Audit has always evolved alongside tools.

From:

- Pencil tick marks
- To spreadsheets
- To cloud workflows
- To AI-assisted testing
- To trained agents
- Autonomous agents

The firms succeeding with AI understand a simple truth: **AI maturity is not a leap. It’s a progression.**

The question isn’t whether your firm will adopt AI.

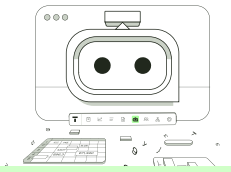
The question is: What level of audit testing readiness have you built, and what must change to move forward?

The 6 Levels of Audit Testing AI Readiness

Level 6

Autonomous Agents

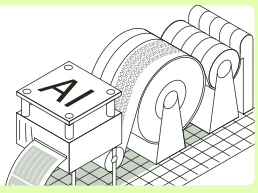
Adaptive agents that sequence and plan procedures dynamically



Level 5

Trained Agents

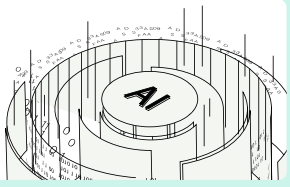
Agents execute approved procedures within defined controls



Level 4

AI-Assisted Testing

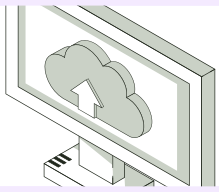
Automation embedded within connected audit workflows



Level 3

Cloud Workflows

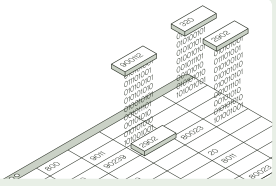
Centralized systems and structured engagement management



Level 2

Spreadsheets and PDFs

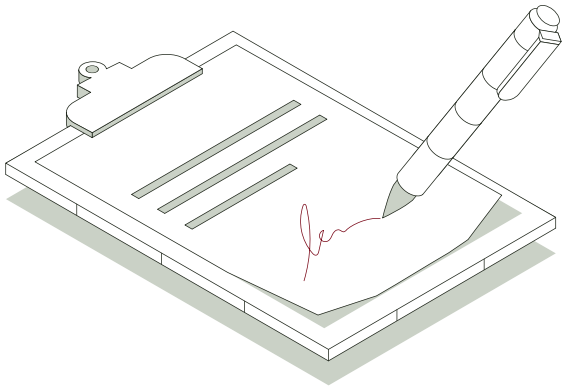
Structured calculations, but disconnected workflows



Level 1

Colored Pencil Tick Marks

Manual testing, sampling, and paper-based evidence



Level 1: Colored Pencils

Manual testing, sampling, and paper-based evidence
The origin of audit testing: binders, printed reports, manual tie-outs, and reviewer memory. Work is human-driven and often offline, with limited system structure.

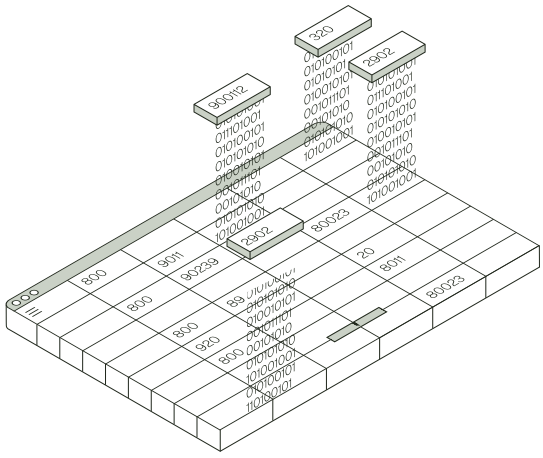
What it looks like

- Heavy sampling
- Printed or PDF workpapers
- Evidence spread across email and shared drives
- Review dependent on individual experience

Risks

- Busy-season bottlenecks
- Limited coverage
- Inconsistent documentation
- Inspection exposure due to gaps in evidence linkage

Focus: Digitize and centralize evidence. Introduce standardized procedure templates and structured workflows.



Level 2: Spreadsheets and PDFs

Structured calculations, but disconnected workflows
Spreadsheets introduced structure and scale — but often became the “system of record” while evidence remained fragmented.

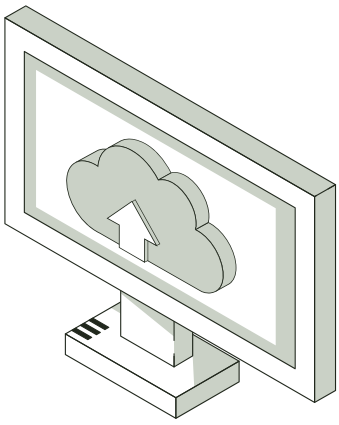
What it looks like

- Excel-based workpapers
- PDFs across systems
- Legacy engagement systems
- Manual data extraction
- Version control challenges
- Evidence stored separately

Risks

- Hidden logic errors
- Reconciliation time drains
- Review friction due to missing support
- Overreliance on spreadsheet discipline

Focus: Standardize data intake. Reduce manual handoffs. Prepare for connected workflows.



Level 3: Cloud Workflows

Centralized systems and structured engagement management
Audit work moves into cloud-based platforms where requests, evidence, and workflows are centralized. Visibility improves — but testing logic may still be largely manual.

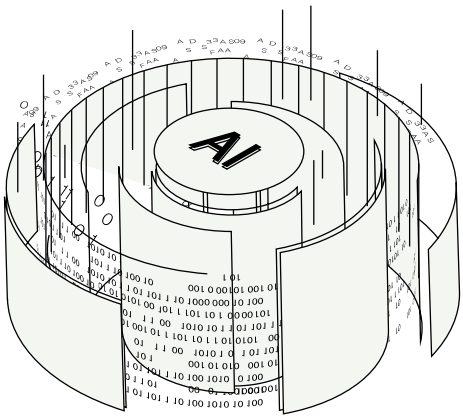
What it looks like

- Centralized PBC portals
- Real-time request tracking
- Shared engagement dashboards
- Standardized templates across teams

Risks

- Automation layered on top of inconsistent data
- Testing logic still manually rebuilt each engagement
- False sense of modernization without scalable testing
- Limited reuse of procedures

Focus: Build reusable, standardized tests. Clean and normalize data pipelines. Make workflows automation-ready.



Level 4: AI-Assisted Testing

Automation embedded within connected audit workflows
AI begins supporting audit execution inside structured systems. Testing expands beyond sampling, and exception identification accelerates — but governance maturity becomes critical.

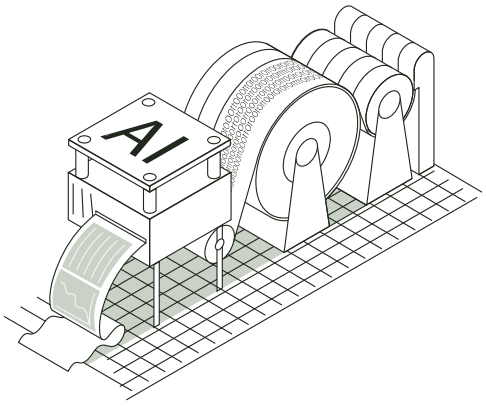
What it looks like

- Full-population matching and recalculation
- Automated anomaly detection
- Exceptions surfaced at scale
- Evidence linked directly to test results

Risks

- Automation outpacing governance
- Difficulty explaining AI logic during inspection
- Efficiency gains measured, quality impact unmeasured
- Inconsistent adoption across engagements
- Exception volume overwhelming reviewers

Focus: Formalize AI governance. Maintain human-in-the-loop review. Measure quality outcomes alongside efficiency. Standardize usage across teams.



Level 5: Trained Agents

Agents execute approved procedures within defined controls
Agentic AI operates within constrained, reviewable boundaries. Agents follow predefined logic and produce audit-ready documentation with human oversight.

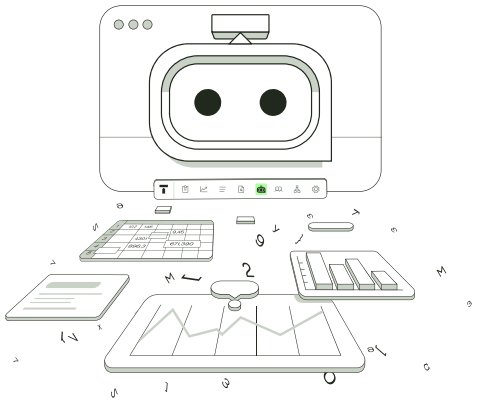
What it looks like

- Agents execute standardized procedures
- Clear audit trail of inputs, logic, and outputs
- Human-in-the-loop approval before reliance
- Controlled rollout by use case

Risks

- Scope creep beyond approved procedures
- Weak change management over agent updates
- Overconfidence in agent outputs
- Inconsistent firm-wide governance

Focus: Constrain agent scope. Embed Auditable AI for fully traceable workflows. Implement formal oversight and change management. Tie deployment to measurable audit quality improvements.



Level 6: Autonomous Agents

Adaptive agents that sequence and plan procedures dynamically
The frontier stage. Agents can adapt workflows, sequence testing steps, and respond dynamically, but always within strict governance and inspection-ready controls.

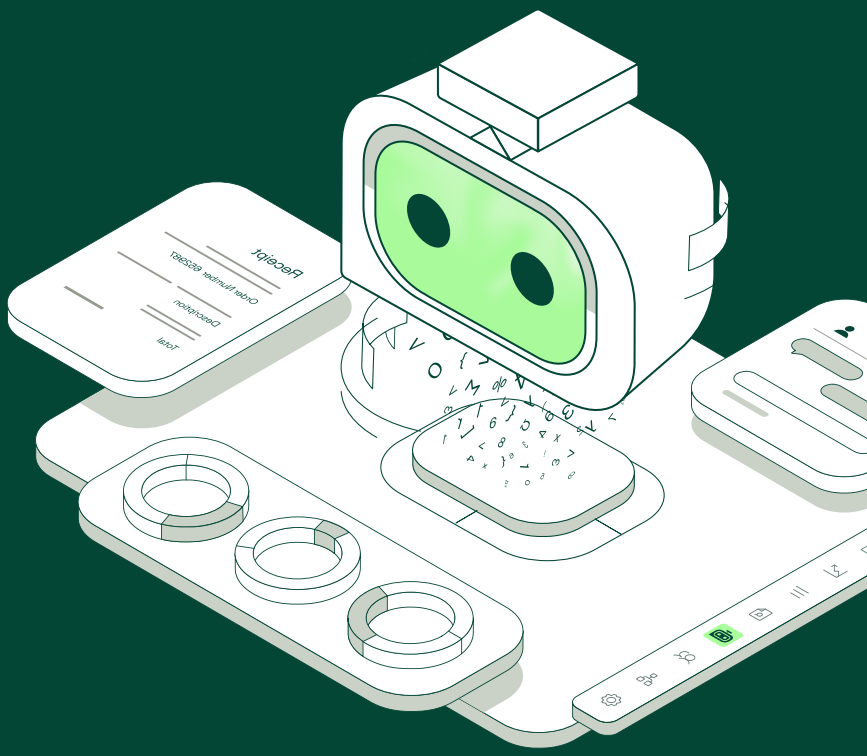
What it looks like

- Agents initiate multi-step procedures
- Dynamic adjustment to data conditions
- Embedded monitoring and observability
- Continuous documentation capture

Risks

- Reduced human visibility into decision sequencing
- Governance frameworks lagging technical capability
- Regulatory uncertainty around autonomy
- High impact if controls fail

Focus: Develop advanced governance and monitoring frameworks. Prove defensibility and reproducibility before expanding autonomy



AI Readiness Assessment Scorecard

Determine Your Audit Testing AI Readiness Level

The path to AI readiness begins with an honest assessment of your current audit testing ecosystem.

This scorecard is designed specifically for audit firms and teams evaluating how prepared they are to adopt AI effectively, from foundational automation through trained agents.

It benchmarks readiness across five core areas:

1. Data and evidence foundations
2. Testing standardization and coverage
3. Workflow connectivity
4. Reviewability and audit trail
5. AI governance and agent preparedness

How to Score

For each question, rate your current capability on a scale of 1–10:

1 (Ad Hoc): Manual, inconsistent, fragmented
5 (Developing): Partial standardization, limited automation, gaps remain
10 (Purpose-Built): Connected, auditable, scalable

Section A: Data + Evidence Foundations

1. Source-of-Truth Data Access: How reliably can your audit team pull complete, standardized datasets (GL, TB, AR/AP, subledgers) without manual file chasing?

1: Manual exports and email uploads; completeness is uncertain
5: Some connectors/templates, but frequent normalization and re-requests
10: Direct ERP integration, repeatable extraction and standardization with minimal rework

2. Evidence Centralization: How centralized and organized is supporting evidence (PBC docs, invoices, bank statements, confirmations) for testing and review?

1: Evidence scattered across email and shared drives
5: Some portal structure, but shadow folders persist
10: PBC portal connects directly with testwork platform

3. Evidence Readiness for Automation: How often is evidence submitted in automation-friendly formats (consistent naming, searchable PDFs, structured exports)?

1: Highly inconsistent; heavy manual cleanup
5: Some standardization; automation works only in limited cases
10: High consistency; evidence is reusable and machine-readable

Section B: Testing Standardization + Coverage

4. Procedure Consistency Across Engagements: How standardized are your core testing procedures across teams and engagements?

1: Everyone performs tests differently
5: Templates exist but are optional; deviations are common
10: Firm-wide standardized templates with governed reuse

5. Sampling vs Full-Population Capability: How often can you test beyond sampling (full-pop matching, automated recalcs) without adding headcount?

1: Sampling is default due to time constraints
5: Full-pop testing used occasionally for narrow procedures
10: Full-pop testing is routine where appropriate, enabled by automation

6. Exception Handling Maturity: How mature is your process for identifying, triaging, documenting, and resolving exceptions?

1: Exceptions handled ad hoc, inconsistently documented
5: Exceptions tracked manually; root causes unclear
10: Systematic exception classification, routing, and documentation

7. Testing Agility: How quickly can you adapt a test when client data structures or evidence formats change?

1: Tests rebuilt from scratch; weeks of effort
5: Some modularity, but still heavy manual rework
10: Configurable workflows; changes handled in hours–days

Section C: Workflow Connectivity

8. Connected Audit Ecosystem: How connected are your systems across data ingestion, evidence, testing, and reporting?

1: Fragmented tools, manual uploads everywhere
5: Partial integrations, but frequent handoffs
10: End-to-end connected workflows with minimal duplication

9. Collaboration + Visibility: How visible is engagement status across requests, submissions, testing, review, and outstanding work?

1: Status lives in spreadsheets and email threads
5: Some dashboards, but multiple sources of truth
10: Real-time engagement visibility across the audit lifecycle

10. Reuse of Testing Work: How much of your testing setup can be reused across periods and engagements?

1: Most work rebuilt each engagement
5: Partial reuse, fragile and inconsistent
10: High reuse with controlled updates and firm-wide scaling

Section D: Reviewability + Audit Trail

11. Traceability to Source Evidence: Can reviewers trace each test result directly back to source documents and data?

1: Review requires chasing evidence across systems
5: Links exist but are incomplete
10: Every result links to supporting evidence with full traceability

12. Documentation Completeness + Timeliness: How consistently does documentation show what was done, by whom, and when?

1: Documentation backfilled late; gaps common
5: Mostly timely, but cleanup required
10: Documentation created in real time with clear approvals

13. IPE and External Evidence Evaluation: How strong is your process for evaluating information produced by the entity (IPE) and other evidence inputs?

1: Assumptions informal; inconsistently validated
5: Some checks exist; limited documentation
10: Systematic evaluation and documented validation

14. Quality Measurement Beyond Efficiency: Do you measure audit quality impact, not just hours saved, when adopting automation?

1: No measurement; success is anecdotal
5: Usage tracked, limited linkage to quality outcomes
10: KPIs tie automation to fewer review notes, stronger coverage, reduced rework

Section E: AI Governance + Agent Readiness

15. Clear AI Use Cases + ROI Discipline: How well-defined are your AI goals, success metrics, and adoption roadmap?

1: AI experimentation everywhere, unclear outcomes
5: A few defined use cases, mixed ROI clarity
10: Prioritized roadmap with measurable benchmarks

16. Human-in-the-Loop Controls: How consistently do you maintain reviewer oversight before AI-assisted outputs become audit evidence?

1: Oversight informal, inconsistent
5: Review checkpoints exist but uneven enforcement
10: Strong human-in-the-loop controls with clear sign-off

17. Transparency of AI Logic: Can auditors explain how AI arrived at outputs (rules, thresholds, matching logic)?

1: Black box results, hard to defend
5: Partial explanation, limited in edge cases
10: Fully explainable logic with review-ready documentation

18. Readiness for Trained Agents: If you introduced trained agents tomorrow, could you constrain them to approved procedures with audit trails and governance?

1: Agents would operate outside controlled procedures
5: Limited pilots possible with heavy supervision
10: Agents can execute defined procedures with full transparency and controlled rollout

Scoring Results: Where Do You Stand?

Add your scores across all 18 questions.

Maximum Score: 180

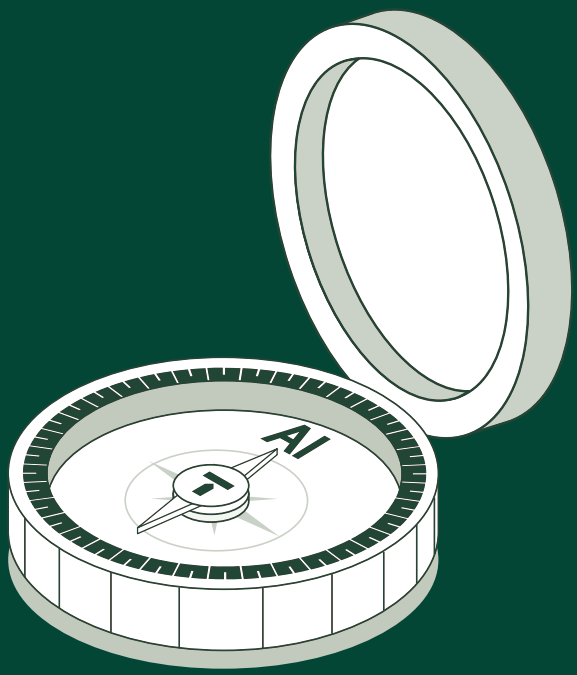
0–60
Foundational stage (Levels 1–2)
Manual testing and spreadsheet-driven workflows dominate

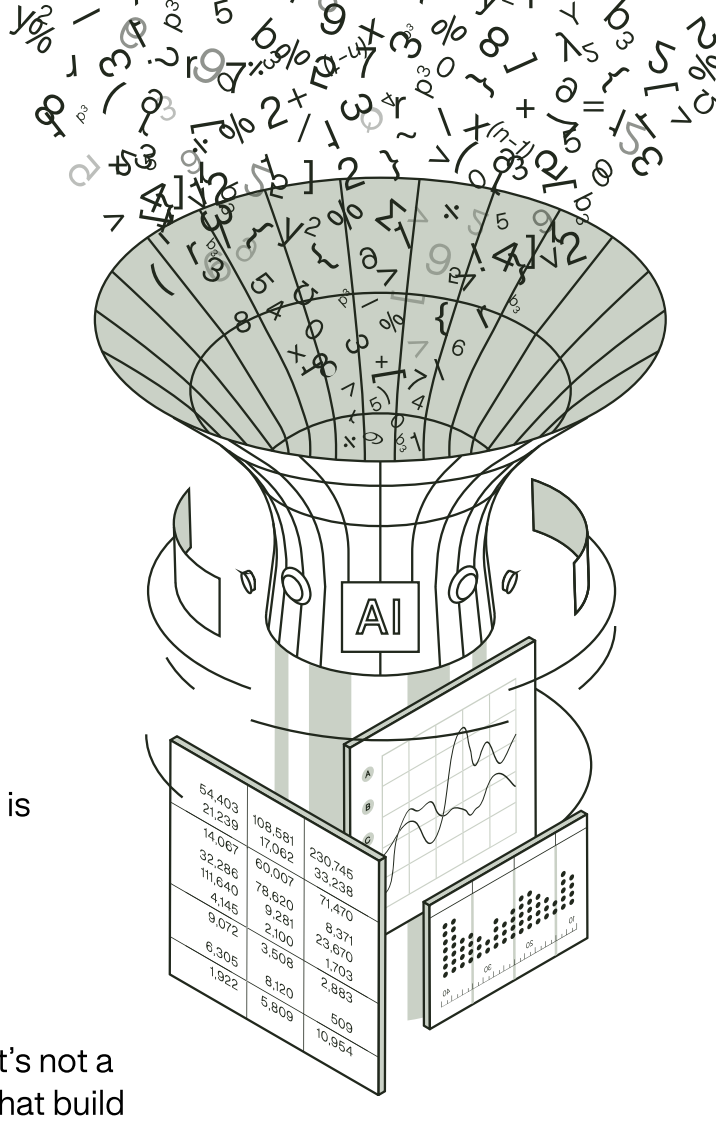
61–95
Developing stage (Level 3)
Evidence is attached but workflows remain fragmented

96–130
Modernizing stage (Level 4)
Connected AI cloud testing is emerging

131–160
Advanced stage (Level 5)
Trained agents operate with governance in place

161–180
Frontier stage (Level 6)
Autonomous agent exploration becomes possible, with strong controls





What to Do Next: Building an AI Roadmap

The most important takeaway from this assessment is simple:

AI readiness is not about being early or late.
It's about being intentional.

If your score indicates foundational work ahead, that's not a setback. It's the smartest place to begin. The firms that build connected, auditable workflows first are the ones best positioned to unlock the full value of AI.

If your score shows strong maturity, your next opportunity is to expand AI from isolated automation into standardized, governed procedures and eventually into trained agents that execute testing with transparency and control.

In every case, the path forward is clearer than ever:

1. Strengthen your data and evidence foundation
2. Standardize repeatable testing procedures
3. Connect workflows into a unified audit environment
4. Measure impact on audit quality, not just hours saved
5. Adopt agentic capabilities only when governance and traceability are ready

This is how audit firms move from experimentation to production impact, avoiding the pitfalls that have caused so many AI initiatives to stall.

Audit firms have an opportunity to do better by building readiness into the process from the start.

Trullion: Turning Readiness Into Reality

AI readiness isn't about jumping to agents.
It's about strengthening the system beneath them.

The firms that scale AI successfully start with connected workflows, clean data foundations, and evidence that is centralized and traceable.

Because in audit, AI must be auditable.

Trullion is built for connected audit — unifying data, evidence, testing, and documentation in one cloud-native workspace. AI isn't layered onto fragmented tools. It's embedded directly into structured, reviewable workflows.

With Trullion, firms can:

- Connect data and evidence into a single source of truth
- Embed AI into audit-ready, traceable procedures
- Expand beyond sampling into scalable testing
- Preserve human oversight and inspection-ready audit trails
- Introduce trained agents within governed, controlled workflows

Because the future of audit isn't just AI. It's connected. Auditable. Agent-ready.

Trusted by industry leaders

Eisai · Taboola · Mitsubishi Chemical · hmd. · Lemonade · SONOS

About Trullion

Trullion is built by former Big Four and top-firm CPAs who've lived the late nights, manual testing, and endless spreadsheets. We help teams move faster, work smarter, and focus on judgment, not busywork.

Ready to move forward with confidence?

For more information: www.trullion.com | info@trullion.com

TRULLION

For more information www.trullion.com info@trullion.com