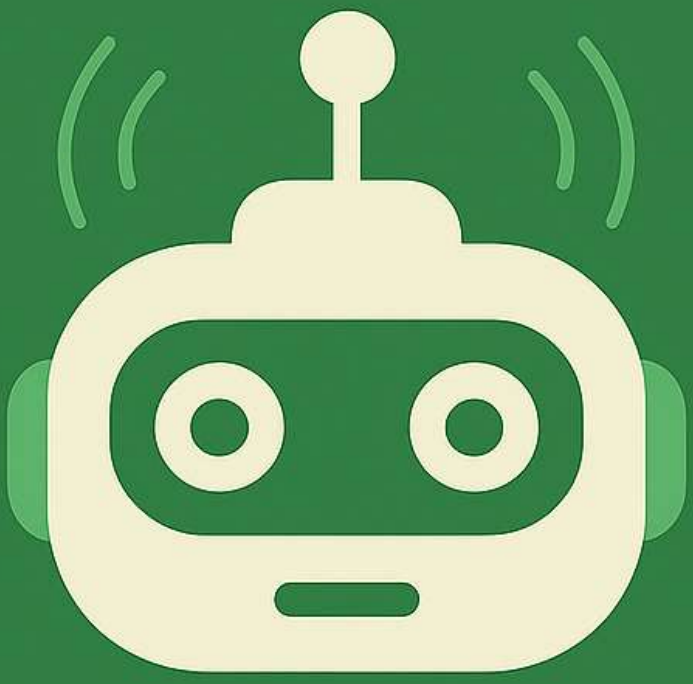
# AI AND AUDITING

## The New Eye Behind the Numbers



01 May 2025


**HAMMAN SCHOONWINKEL**

*With some assistance from a very helpful robot*

# Introduction

Artificial Intelligence (AI) is no longer a distant frontier in auditing — it is reshaping the profession from within. Modern auditors are expected not only to understand AI tools, but to collaborate with them: from machine learning models that score risk across millions of journal entries, to natural language processors that digest board minutes and flag governance concerns. Yet while these technologies promise unprecedented speed and scale, they also raise critical questions about judgment, reliability, and accountability.

This report explores the role of AI at each stage of the audit lifecycle — from planning and risk assessment, through tests of controls and substantive procedures, to audit reporting and completion. It further examines how the client's own use of AI alters the control environment and influences audit strategy, requiring auditors to evaluate not just human processes, but algorithms.

This work adopts a phase-by-phase structure aligned to the International Standards on Auditing (ISAs), with focused discussions on how AI tools support specific audit objectives — such as evaluating the reasonableness of accounting estimates, identifying high-risk outliers, or checking the completeness of disclosures. Dedicated sections address professional skepticism in the age of AI, as well as the emergent audit tools being developed by global firms like Deloitte (Argus), PwC (GL.ai), and KPMG (Clara).

But the report is not uncritical. It repeatedly emphasizes the limitations of AI, especially around explainability, bias, over-reliance, and documentation responsibilities. The inclusion of black box models within a profession built on transparency and defensibility creates a tension that cannot be ignored. AI may excel at pattern recognition — but auditing demands justification.

Designed for readers who already understand traditional audit procedures, this report goes beyond technical application to provoke deeper reflection: What does it mean to audit in an age where the tools may know more than

we do, but cannot explain how? AI does not replace the principles of auditing — but it does force us to reinterpret them, in a landscape that is expanding faster than any checklist can contain.

---

# 1. Planning and Risk Assessment

The audit lifecycle begins with planning and risk assessment. In this phase, auditors gain an understanding of the client's business and identify areas of higher risk of material misstatement. AI can dramatically enhance risk assessment procedures by processing large volumes of data – both structured and unstructured – to pinpoint anomalies and risk indicators that might be missed by traditional sampling techniques.

## 1.1 NLP for Understanding the Entity and Its Environment

According to ISA 315 (Revised 2019), the auditor is required to "obtain an understanding of the entity and its environment, the applicable financial reporting framework, and the entity's system of internal control..." (ISA 315.16). This understanding is foundational for identifying and assessing the risks of material misstatement and for tailoring appropriate audit responses. As outlined in ISA 315 (Revised 2019), auditors are required to develop a comprehensive view of the entity and its environment — including factors such as industry conditions, regulatory landscape, and external economic forces; the nature of the entity's operations and revenue streams; ownership and governance structures; major investment and financing decisions; and the entity's strategies, objectives, and business risks. Additionally, auditors must gain insight into the components of the internal control system, particularly those relevant to financial reporting.

Critically, much of this information is qualitative in nature. Rather than being found in ledgers or trial balances, it is embedded in narrative sources — such as board and committee meeting minutes, internal audit reports, policy manuals, and risk registers. In certain engagements, particularly those involving fraud risk or forensic elements, this may also extend to

whistleblower reports, contracts, and internal communications. These unstructured sources often contain the subtext and early indicators of risk — including signs of internal disagreements, tone-at-the-top concerns, or operational challenges — that do not surface in numerical summaries.

Traditionally, auditors manually skimmed these unstructured documents for red flags — a process both time-consuming and prone to oversight. Natural Language Processing (NLP), a field of AI, enables auditors to review entire corpora of such documents efficiently, improving both the breadth and depth of risk identification.

**How NLP Works in Practice**

Modern NLP tools do more than scan for specific keywords. They interpret context and meaning, allowing for semantic understanding that mirrors how a human might infer risk — but at massive scale.

Here's how a typical NLP-assisted audit procedure might unfold:

1. Ingest Documents: The auditor uploads hundreds or thousands of documents (minutes, reports, policies) into the system.
2. Preprocessing: The AI corrects OCR errors, cleans formatting, segments text into meaningful paragraphs and sentences.
3. Contextual Risk Parsing: Instead of just matching words like "fraud" or "override," modern NLP tools use semantic vector embeddings (such as BERT or GPT-derived models) to assess meaning. For example, the AI might recognize that the phrase:

   *"Management remains cautious about the company's liquidity as receivables continue to lag and supplier obligations mount"*

   is semantically equivalent to a concern about cash flow risk, even though none of the traditional keywords appear.
4. Keyword + Semantic Matching: Tools use curated risk lexicons (e.g., "unrecorded liability," "non-compliance," "restatement," "over-

budget") but extend these through context-aware models to flag euphemisms or indirect risk indicators. For instance:

- "System upgrades have been delayed again due to vendor issues" (suggesting potential project mismanagement)
- "Several key controls are pending automation" (suggesting gaps in operational effectiveness)

5. Thematic Clustering & Sentiment Analysis: AI identifies patterns across documents. If multiple reports contain growing negative sentiment or recurring mentions of unresolved issues, the NLP model can highlight these as emerging themes. For instance:
    - Clusters of paragraphs discussing IT instability or cybersecurity concern.
    - Repeated references to strained working capital.
6. Output & Summary: The AI produces a report summarizing key findings. This may include flagged sentences, thematic heatmaps (e.g., which sections had most compliance-related language), and extracted entities (names, dates, roles, etc.).

**ISA Alignment and Examples**

NLP supports numerous ISA 315 mandates:

- ISA 315 requires consideration of the tone at the top, control environment, and management's attitude toward risk. These are often captured in informal language ("we took a risk on this deal", "unforeseen exposure", "pressured to close books early") that NLP can pick up.
- ISA 315 requires the auditor to determine whether the understanding obtained is sufficient. NLP helps ensure completeness by reviewing all available documents, not just those selected for manual review.
- ISA 240 mandates the identification of fraud risk factors. NLP can reveal red flags such as management override, poor segregation of

duties, or pressure on financial reporting timelines through indirect language.

For example, an NLP tool used on board minutes and internal audit summaries flagged frequent discussions around:

- Delays in ERP integration
- Vendor payment renegotiations
- Executive team restructuring

None of these were formally presented as "risks" to the auditors — yet when reviewed together, they signalled potential issues in revenue recognition (ERP delay), liquidity (vendor renegotiation), and governance (executive churn).

**Key Considerations for Audit Documentation**

Auditors must still comply with ISA 230 (Audit Documentation) when using AI. Key aspects include documenting:

- The documents reviewed and their source
- The NLP tool used (including version)
- The risk lexicon and semantic thresholds applied
- A summary of flagged findings and how they were responded to (discussed with management, inspected further, tested, etc.)

ISA 500 also remains relevant: the sufficiency and appropriateness of evidence from AI tools depends on their accuracy and the completeness of source material. If certain documents were missing from the NLP input (e.g., a missing risk register or draft policy), auditors must acknowledge that limitation.

## 1.2 Preliminary Analytical Procedures

According In accordance with ISA 315 (Revised 2019), auditors are required to perform risk assessment procedures to obtain a sufficient understanding of the entity and its environment — including its financial performance and position — for the purpose of identifying and assessing risks of material misstatement. One such procedure is the use of analytical procedures

during the planning phase, which, although not governed by ISA 520, is discussed in detail in ISA 315.A27–A31 as a tool for detecting anomalies, inconsistencies, and trends that may point to elevated risk.

These preliminary analytical procedures are typically high-level comparisons of financial and non-financial data, aimed at uncovering unusual or unexpected relationships. ISA 315.A27 highlights their role in identifying "inconsistencies, unusual transactions or events, and amounts, ratios, and trends that indicate matters that may have audit implications," including potential fraud risk factors. As such, these procedures serve as a key input into the auditor's inherent risk assessment, especially where the entity operates in volatile, regulated, or rapidly changing environments.

**How Analytical Procedures Inform Risk Assessment**

As explained in ISA 315.A28–A29, these procedures may incorporate both financial and non-financial information — for example, comparing sales revenue to the square footage of retail space, or reviewing gross margin trends in light of known supplier price increases. The relationships observed can help the auditor:

- Identify unexpected deviations from historical patterns;
- Detect non-linear or erratic changes across accounts that may suggest misstatement;
- Surface correlations that no longer hold true, indicating business changes or potential manipulation.

Such observations provide an "initial broad indication," as per A29, of where risks may be concentrated and where more specific audit procedures might be necessary.

**AI-Augmented Preliminary Analytics**

While traditional approaches rely on side-by-side comparison of current versus prior period data or against budgets, modern audit teams increasingly use AI-driven analytics to transform this process from a

checklist into a predictive, insight-generating activity. The use of automated tools and techniques, described in ISA 315.A31, reflects this evolution.

AI-based tools support auditors by:

1. Learning historical data patterns over multiple periods to establish a more statistically robust baseline;
2. Comparing actuals to dynamic expectations, not just budgets, but models built from macroeconomic trends, industry benchmarks, or operational metrics;
3. Detecting complex patterns, such as year-end adjustments clustering in certain accounts, or unexpected changes in the correlation between revenue and receivables;
4. Visualizing large datasets to make anomalies more apparent — for example, highlighting operating expense trends that deviate significantly from inflation-adjusted historical norms;
5. Ingesting non-financial data, such as staff turnover, production capacity, or web traffic, and linking those to performance indicators (e.g., identifying whether declining conversion rates match revenue dips).

For instance, an AI system may note that although a client's top-line revenue increased by 10%, website traffic fell by 20%, customer returns rose sharply, and social sentiment around a product line declined. These non-financial anomalies suggest that the reported revenue might deserve additional scrutiny — perhaps indicating premature recognition, changes in sales terms, or errors in cut-off.

**Informing the Audit Plan**

The outputs of these AI-enhanced analytical procedures help the auditor to:

- Focus more precisely on classes of transactions or balances that show unexpected behaviour;

- Adjust materiality and sampling thresholds in areas where variability or volatility is higher than expected;
- Identify potential risk of fraud as outlined in ISA 240 (e.g., unusual revenue recognition patterns);
- Determine whether certain assertions (like completeness or valuation) are inherently riskier.

These procedures also support the identification of significant classes of transactions, account balances, and disclosures (SCOTABDs) under ISA 315.27–28, where preliminary anomalies point to the need for deeper understanding or more detailed testing.

**Documentation Responsibilities**

As required under ISA 230, auditors must document:

- The analytical procedures performed;
- The nature of expectations developed (e.g., via AI forecasting models or historical baselines);
- Any unusual or inconsistent relationships identified;
- Their assessment of whether these relationships indicate potential risks of material misstatement, and how those risks were addressed.

If automated tools are used (as acknowledged in ISA 315.A31), the documentation should include:

- The nature and source of the data used;
- The functionality of the tool (e.g., whether it used machine learning, regression, or visualization);
- A summary of flagged items or patterns;
- Auditor commentary on how these informed the broader risk assessment.

## 1.3 AI-Driven Transaction-Level Anomaly Detection

In addition to the high-level analytical procedures described in Section 1.2, auditors are increasingly using AI-driven anomaly detection to analyse data at the transaction level. While this also qualifies as an analytical procedure under ISA 315 (Revised 2019), its methodology and depth of focus differ significantly. Traditional analytical procedures typically assess aggregated relationships (e.g., gross margin trends, receivables turnover) to spot unusual fluctuations, as envisioned by ISA 315.A27–A30. By contrast, anomaly detection technologies — particularly those powered by machine learning — scan entire populations of transactional data, such as journal entries or invoice records, to identify items that deviate from learned expectations or established patterns.

This makes anomaly detection a specialized subclass of analytical procedures: one that works from the bottom up rather than top down. ISA 315.A31 specifically acknowledges the growing use of automated tools and techniques — often referred to as "data analytics" — which allow auditors to apply analytical procedures in more advanced ways. Anomaly detection fits squarely within this evolution.

**From Sample-Based Mindsets to Full-Population Analysis**

Historically, auditors assessed risk using selective document reviews, interviews, and basic comparisons — rarely analysing all transactions prior to fieldwork. But AI tools like MindBridge AI Auditor or PwC's GL.ai allow practitioners to ingest full ledgers, apply scoring algorithms, and instantly flag unusual items. These anomalies might include:

- Journal entries posted outside of working hours
- Transactions with rounded, high-dollar values
- Entries involving unexpected account pairings (e.g., debiting "inventory" while crediting "director loan")
- Frequent use of manual overrides
- Use of non-routine adjustment codes (e.g., JEs marked "correction" or "urgent")

Such results are particularly relevant to the auditor's identification of risks of material misstatement under ISA 315.12 and .19, and directly support paragraph A28, which notes that analytical procedures may assist in understanding how inherent risk factors (such as complexity, subjectivity, or change) affect susceptibility to misstatement.

Moreover, AI systems do more than just apply fixed rules. They use unsupervised machine learning to establish what "normal" looks like for a given client (based on historical patterns, account relationships, user behavior, etc.) and then flag items that statistically deviate. This form of context-aware analysis moves beyond traditional red-flag logic and allows auditors to surface unexpected or subtle risks — not just confirm known risk indicators.

**Real-World Impact on Planning**

This kind of granular anomaly detection supports more informed and targeted audit planning:

- It narrows down high-risk transactions or accounts requiring more robust audit procedures.
- It helps auditors determine whether particular assertions (e.g., accuracy, occurrence, cutoff) deserve additional scrutiny.
- It provides a data-driven basis for identifying potential fraud risks under ISA 240, especially those related to management override (ISA 240.32).
- It aligns with ISA 315.A49's emphasis on examining journal entries and other adjustments as part of understanding internal control and the risk of fraud.

AI-based anomaly detection does not replace the auditor's professional skepticism or planning judgment. Rather, it extends the auditor's visibility into areas that would have otherwise been examined only post hoc or via limited sampling.

**Practical Note**

Some audit firms integrate anomaly detection directly into the initial risk assessment dashboard, enabling engagement teams to visualize concentrations of anomalies across time periods, business units, or user profiles. For example, Deloitte's Cortex AI tool maps anomaly clusters visually, helping teams prioritize walkthroughs or inquiries. In this way, anomaly detection becomes more than a technique — it is a risk radar system that enhances ISA 315-compliant planning with broader coverage and earlier insights.

## 1.4 Pattern Recognition and Risk Scoring

Were traditional analytical procedures (Section 1.2) focus on high-level metrics and anomaly detection (Section 1.3) isolates individual outliers within large datasets, modern AI techniques now go further — identifying complex patterns across multiple dimensions and assigning risk scores that prioritize audit attention based on learned correlations. These systems are capable of inferring elevated risk not from a single red flag, but from a constellation of otherwise subtle indicators.

This development aligns closely with the auditor's responsibility under ISA 315 (Revised 2019), paragraph 19, to obtain a sufficient understanding of the entity and its environment to identify and assess the risks of material misstatement, whether due to fraud or error. In particular, ISA 315.A28 encourages auditors to consider how combinations of inherent risk factors — such as change, complexity, and subjectivity — may increase susceptibility to misstatement. Pattern recognition models operationalize this idea.

**From Anomalies to Risk Profiles**

A full-population anomaly detection tool might flag a transaction because it was posted at 2:00 a.m. by a junior employee. In contrast, a pattern recognition model might surface a more complex concern: that during the final week of the reporting period, multiple users in one department posted large adjusting entries, all just under manual review thresholds, and each associated with newly created vendors. None of these factors, alone, would

trigger an anomaly detection alert — but together, they tell a compelling story.

This shift reflects a movement from rule-based alerting (e.g., "flag all journal entries over R1 million") to data-driven inference: using supervised machine learning to detect patterns historically associated with misstatements. These models are often trained on labeled data — past audit findings, fraud case databases, or internal control failure logs — to develop an understanding of what combinations of activity have correlated with risk in the past.

These models are not static. They learn, evolve, and adapt — helping auditors continuously refine their risk lens as new data becomes available.

**How It Works**

A typical AI risk scoring engine applied in the planning phase may:

- Integrate transaction-level data, master data, and user behaviour logs.
- Engineer hundreds of features (e.g., average transaction size by user, frequency of reversals, change in vendor master file activity, etc.).
- Use machine learning algorithms — such as random forests, gradient boosting, or neural networks — to generate a risk probability score for each transaction, account, or business unit.
- Visualize the results in a heatmap, dashboard, or entity profile for the auditor.

Unlike anomaly detection (which may simply surface a strange timestamp), risk scoring models look at patterns over time, across multiple variables, and in specific contexts. Some systems even assign confidence levels or explainability metrics to indicate how much weight specific variables carried in the risk prediction.

Artificial Intelligence, particularly through natural language processing, anomaly detection, and predictive modelling, is redefining how auditors fulfil

their risk assessment responsibilities under ISA 315 (Revised 2019). Rather than replacing auditor judgment, these tools enhance the auditor's ability to obtain a deep and broad understanding of the entity and its environment — from identifying implicit risk cues in qualitative documentation to uncovering subtle patterns in transactional data that may otherwise go unnoticed.

By augmenting traditional preliminary analytical procedures and enabling full-population analysis, AI acts as a force multiplier during the planning phase. It allows auditors to go beyond narrow sampling and surface insights at both the micro (transactional) and macro (entity-level) scale. This aligns directly with the ISA's call for auditors to move beyond checklists and toward a more dynamic, evidence-informed understanding of risk.

---

# 2. Test of Controls

In the audit of financial statements, tests of controls (ToC) provide the auditor with evidence about whether key internal controls — previously assessed as well-designed and implemented during Risk Assessment Procedures (RAP) — have operated effectively throughout the period. As outlined in ISA 330, these procedures enable the auditor to determine whether reliance can be placed on controls when designing the nature, timing, and extent of substantive procedures.

Traditionally, ToC has drawn on four core methods: inspection, observation, inquiry, and reperformance. Each of these techniques is aimed at evaluating, not merely the existence or design of a control, but its consistent application and effectiveness across relevant transactions or events. Importantly, ToC is distinct from RAPs, which focus on understanding and identifying risks. In contrast, ToC is confirmatory — it asks not whether a control exists or appears sound, but whether it was actually performed, by the right person, at the right time, in the right way.

This distinction has significant implications for the use of technology — and especially artificial intelligence (AI). In the risk assessment phase, AI's strength in processing large volumes of unstructured data, surfacing red flags, and performing high-level semantic analysis often leads to clear efficiency and insight gains. In ToC, however, the audit context shifts. The sample sizes are smaller, the procedures more rule-based or binary, and the questions more narrowly scoped. The practical benefit of AI thus varies greatly across techniques, depending on the digital maturity of the client, the nature of the control evidence, and the complexity of the judgment involved.

This chapter critically examines the use of AI in each of the ToC procedures — inspection, observation, inquiry, and reperformance — highlighting where technology offers genuine efficiency or enhancement, where it introduces friction or overengineering, and where human judgment might remain irreplaceable. It rejects both techno-optimism and blanket conservatism. Instead, it aims to show that AI's role in ToC is context-specific: a function of control structure, data accessibility, and audit objectives. By scrutinizing both the strengths and the constraints of AI within each ToC method, we better understand how to deploy these tools strategically — and where to preserve manual engagement as a matter of both audit quality and professional skepticism.

## 2.1 Inspection

In traditional audit methodology, inspection refers to the auditor's examination of records, documents, or tangible assets to gather evidence about the operation of internal controls. According to ISA 500, Audit Evidence, inspection includes reviewing both internally generated documents (such as policies, control narratives, or system logs) and externally sourced materials (such as contracts or third-party confirmations). Within a test of controls, inspection is a formal procedure used to evaluate whether controls previously assessed as appropriately designed and implemented (during the risk assessment procedures under ISA 315) are, in fact, operating effectively throughout the period — for

instance, by verifying that approvals, reconciliations, or supervisory reviews were actually performed and documented.

**Where AI Adds Value to Inspection**

AI enhances the traditional inspection process by making it feasible to analyse entire populations of control-related documents — not just samples — where such data is available. This includes structured logs as well as unstructured text.

Examples of AI-enabled inspection include:

- Verifying reconciliation checklists: AI systems can scan a full set of monthly reconciliations to check if each was signed and dated by the correct personnel.
- Detecting missing or delayed approvals: AI can flag any documents where an approval signature is missing, where the approval occurred after the transaction date, or where required fields are incomplete.
- Extracting review notes and annotations: NLP can locate and summarize control annotations or exception notes across large volumes of compliance or internal audit reports.
- Metadata review: AI tools can detect inconsistent metadata (e.g., conflicting time stamps or authorship fields) in approval documents, which may suggest irregularities.
- Cross-department consistency: AI can compare documentation across divisions to assess whether the same control is applied uniformly, or whether certain areas are falling behind in documentation practices.

These capabilities extend the auditor's reach far beyond traditional manual inspection, which is often limited to a handful of samples. Especially in large, decentralized, or highly automated environments, AI makes it possible to inspect for control execution at scale.

**Where AI Is Less Suitable for Inspection in ToC**

Despite its promise, AI is not always the right tool for test of controls inspection. To understand why, it helps to contrast this phase with risk assessment procedures (RAP).

AI is extremely well-suited to the RAP phase, particularly in inspection-based procedures like reviewing 100% of the general ledger for anomalies, parsing all board minutes or internal audit reports for red flags and scanning narratives to build understanding of tone, culture and design of controls. AI works well because you are dealing with high-volume, unstructured and non-standardized data. You want AI to do risk triage – to surface issues a human would likely miss due to fatigue or bias. You are not trying to conclude whether a control was done properly, but rather where you should look deeper.

In Test of Controls (ToC), the AI-powered inspection model does not always translate well. Firstly, the volume is lower, as you are usually testing a sample, maybe 20-60 items per control. Secondly, the procedure is typically binary or checklist-based, as you check for example for the presence of a signature, whether it was dated within 5 working days, or whether the signatory is authorized. This is not where AI shines. The benefit of automation diminishes when the task is simple, and the volume is small.

In RAP, the human cost of reading 2 000 board minutes is high, so here AI saves tremendous time. In ToC, the human cost of inspecting 30 reports is low, so here AI may introduce more friction (e.g. digitizing and uploading the sample, setting parameters, validation AI outputs).

AI's value proposition in ToC Inspection is limited by:

1. Digital Maturity of the Client: Unless the client has a fully digital document environment (with reliable metadata), feeding AI with structured samples is logistically burdensome.
2. Nature of the Control Evidence: Approval signatures and timestamps are low-complexity tasks, so AI may be overkill.
3. The Limited Need for Pattern Recognition: In ToC, the question isn't "where are the anomalies?" but "did this specific control operate as

expected?" across a targeted, defined sample. There's little need for AI's semantic or clustering capabilities — which is exactly where its strength lies.

**System-Oriented CAATs and AI Inspection**

AI-enhanced inspection builds upon traditional system-oriented CAATs (Computer-Assisted Audit Techniques) previously taught in audit curricula.

*Code Analysis*

Traditionally, code analysis involved human specialists reviewing the source code of custom-built systems to verify that automated controls were correctly programmed. This is most relevant where off-the-shelf assumptions about functionality cannot be made.

Modern AI tools offer a conceptual parallel. While they cannot truly analyse raw code reliably, they can assist in interpreting configuration files, flagging unusual logic patterns, or even explaining approval rules embedded in systems. In this way, AI may act as a first-pass filter — highlighting areas that warrant deeper expert review — but does not yet replace the technical expertise required for full source code analysis.

*SCARF (System Control Audit Review File)*

SCARF refers to a legacy technique in which an audit module is embedded into a client's system to monitor control activities in real time. It was groundbreaking for its era but is now largely seen as outdated due to system performance issues, limited adaptability, and incompatibility with modern cloud environments. Instead of embedding audit code directly into live systems, modern auditors use AI-enabled log analysis tools to retrospectively analyse activity logs and control events. These tools review digital footprints — such as who approved a transaction, when it occurred, and whether any procedural step was bypassed — to assess control operation without interfering with the client's production environment.

Both traditional CAATs and AI-based inspection techniques serve a similar purpose: to gather persuasive audit evidence that controls are not only

theoretically present but are operating in practice. The key distinction lies in how that evidence is gathered — AI enables broader, more scalable, and more intelligent document and system analysis without the same level of system intrusion.

In this way, AI-enabled inspection tools represent the next generation of audit technologies, building on past CAAT methodologies while offering far greater scalability and efficiency. They empower auditors to inspect control documentation comprehensively, ensure alignment across departments, and flag gaps that may indicate weaknesses in control design — all foundational steps in an effective test of controls.

In summary, AI has introduced powerful enhancements to the inspection phase of testing controls — but it is not a magic bullet. Its advantages are strongest in high-volume, low-structure contexts like risk assessment. In ToC, especially when dealing with straightforward, low-volume procedures, the return on AI investment diminishes. Nonetheless, in highly digital environments or where document metadata and consistency are critical, AI-enabled inspection tools remain a valuable part of the modern audit toolkit.

## 2.2 Observation

Observation, as one of the core test of controls techniques outlined in ISA 330.A29, refers to the auditor's direct witnessing of control activities in action. Unlike inspection (which deals with documentary evidence of past events), observation is concerned with real-time confirmation that controls are performed as designed — especially when those controls do not leave behind a durable audit trail. Though ISA 330 does not define observation in great detail, auditing standards and literature agree that observation is most suitable for evaluating manual controls or behavioural elements that are difficult to prove retrospectively. An example of an observation.

Examples of observation-based tests of controls include:

- Verify whether access control procedures are actually followed – not just documented.

- Watching a cashier count cash at the beginning and end of day to verify the control over cash reconciliation.
- Attending a physical inventory count to ensure counting instructions are followed and properly supervised.
- Observing the steps taken during a journal entry process to confirm that preparer and approver roles are functionally segregated.

In each of these scenarios, the auditor is not relying on paperwork after the fact, but directly witnessing the event. This immediacy is what gives observation its evidentiary strength — especially where the entity's documentation is weak or where behaviour and adherence to process are in question.

**AI's Limited Role in Observation**

AI excels in domains like document analysis, pattern recognition, and anomaly detection across large datasets — but its value sharply declines when it comes to real-time observation of human behaviour. Observation is inherently judgmental and context-sensitive. It involves noticing body language, hesitations, procedural shortcuts, or environmental factors — areas where current AI lacks the interpretive nuance of a human auditor.

In this sense, AI does not replace observation. It cannot attend a stock count, walk through a building, or assess whether the tone of a manual procedure is compliant or performative. Instead, it may offer limited, supplementary support in highly digital or remote contexts — for example, where physical access is restricted.

**Using Video Footage or Livestreams**

During the COVID-19 pandemic and in geographically remote audits, firms began using livestreams, CCTV feeds, or recorded video clips to conduct observation-style procedures. These include:

- Watching a live video feed of an inventory count.
- Reviewing CCTV footage of access control enforcement at sensitive locations.

- Using recorded sessions of cash handling or reconciliation procedures.

The distinction between observation and inspection becomes blurred when using video. If the auditor watches a livestream, this can still qualify as observation, preserving the immediacy and involvement required. However, if the footage is reviewed after the fact, the procedure would more appropriately be classified as inspection. In practice, this classification is not doctrinally critical — what matters is whether the procedure provides sufficient appropriate audit evidence. ISA 500 requires the auditor to evaluate the reliability of audit evidence, including whether video footage is complete, unaltered, time-stamped, and clearly linked to the control being tested.

**AI Embedded Within Video Evidence**

While the video itself is not AI, artificial intelligence can be layered onto video streams to enhance the observation process. AI particularly adds value when the scale, complexity, or continuity of observation exceeds human capacity. A strong example of AI-driven observation in audit is livestock inventory verification. In large-scale farming operations, manually counting livestock during a stock count may be impractical. Drone footage, paired with AI-powered object recognition, can count moving animals with accuracy and differentiate between species, sizes, or tagged vs. untagged animals.

This elevates the audit procedure from traditional observation to an AI-enhanced inspection of a real-time process — providing stronger assurance with broader coverage.

However, not all use cases justify AI enhancement. If the task is simple — for example, confirming a supervisor observes a daily till count — then live presence or even video alone may suffice. In such low-complexity scenarios, AI may add cost or complication without proportionate benefit.

In summary, AI currently has a limited but growing role in audit observation. While it cannot replace the human act of witnessing controls in action, it can

extend observational reach in certain cases, especially via video analysis. However, many traditional observation tasks remain better suited to human auditors — underscoring the continued importance of judgment, presence, and contextual awareness in tests of controls.

## 2.3 Inquiry

Inquiry is a fundamental audit procedure defined in ISA 500 as seeking information from knowledgeable individuals, either within or outside the entity. It is one of the core techniques listed alongside inspection, observation, reperformance, and analytical procedures. However, both ISA 500 and ISA 330 emphasize that inquiry alone is not sufficient to obtain persuasive audit evidence when testing the operating effectiveness of controls.

**When Inquiry Is Useful in ToC**

Although not sufficient on its own, inquiry can provide value in three key ways during tests of controls:

1. Clarifying Execution Details: Inquiry may help the auditor understand how a control was executed in cases where documentation is ambiguous. For example, after inspecting a signed approval, the auditor might inquire who reviewed the supporting documents and how exceptions were handled.
2. Explaining Deviations or Gaps: If the auditor finds an instance where a control appears not to have operated (e.g., a missing approval), inquiry can clarify whether this was an oversight, a known exception, or an undocumented change in process.
3. Evaluating Consistency Across Roles: Inquiry can be used to corroborate information across departments, identifying inconsistencies in how a control is described or applied. Discrepancies may trigger further testing.

**Appropriateness of Inquiry When No Documentation Exists**

A common but flawed assumption is that inquiry is particularly helpful in situations where no documentation exists. This assumption is problematic for two main reasons:

1. Insufficient Audit Evidence: According to ISA 330.A26, inquiry alone must be corroborated by other audit evidence. If no documentation or alternative evidence is available, the auditor cannot sufficiently test the operating effectiveness of the control. Consequently, reliance on such a control is inappropriate.

2. Circular Logic Risk: If a control lacks documentation, it raises questions about whether the control was properly designed and implemented, as required by ISA 315. Without evidence of design and implementation, the auditor is unlikely to perform tests of controls (ToC) on that control at all.

That said, there are niche scenarios-particularly with entity-level controls-where documentation may be sparse, but inquiry combined with observation or other corroborative evidence can still provide useful insights. Examples include:

- Small Entities: In smaller organizations with informal oversight, the auditor may observe the owner's involvement and corroborate inquiries with other audit evidence.
- Organizational Culture and Ethics Controls: Controls related to tone at the top or ethical culture may be assessed indirectly by combining inquiry with observations of management behaviour and other indicators.

**Limitations of Inquiry**

ISA 500 and ISA 330 are clear that inquiry on its own is weak evidence:

- It is subjective and reliant on the truthfulness and knowledge of the respondent.
- It does not produce durable, reviewable audit evidence unless corroborated by other means.

- It may be impacted by bias, misinterpretation, or intentional misstatement.

Therefore, auditors are required to design their audit procedures such that inquiry supports — rather than substitutes — more objective forms of testing.

**AI's Role in Enhancing Inquiry**

While AI cannot perform a human interview or interpret tone in a conversation the way an auditor can, it does support inquiry procedures in innovative ways:

- Summarizing Narrative Responses: In large-scale internal control surveys or questionnaires, AI can analyse open-ended responses from dozens or hundreds of employees. Natural Language Processing (NLP) can identify recurring themes, contradictions, or sentiment shifts across departments, allowing auditors to focus follow-up inquiries more effectively.
- Identifying Inconsistencies: AI tools can compare responses from multiple sources (e.g., between compliance, finance, and operations departments) to detect misalignments in how controls are described or understood.
- Highlighting Risky Language: AI models trained on control failure scenarios can flag language patterns that suggest weak control environments — such as vague responsibility assignments or justifications for overrides.
- Pre-screening Internal Communication: Where appropriate and permitted, AI can scan internal chat or email transcripts to find deviations from policy or tone — for instance, suggesting that a control was circumvented. While not a replacement for direct inquiry, this can serve as an input into the auditor's questioning.

**Caution in Using AI in Inquiry**

While AI enhances the efficiency of inquiry — such as analysing narrative responses or highlighting inconsistencies across departments — it cannot

replace the auditor's professional judgment. Critical aspects of inquiry still require human skill, including:

- Selecting the appropriate individuals to question.
- Interpreting tone, hesitation, and non-verbal cues.
- Recognizing when a response is overly rehearsed, evasive, or deliberately misleading.

Moreover, AI systems struggle with subtle forms of communication such as sarcasm, cultural nuance, or emotional manipulation — particularly in sensitive or high-stakes contexts. These are precisely the scenarios where professional skepticism becomes most important. As discussed and debated in Section 6, the auditor's ability to maintain a questioning mindset, assess credibility, and respond to uncertainty with critical analysis might not yet be able to be replicated by machines.

In summary, AI can support inquiry as a strategic aid, helping auditors ask better questions and process responses more effectively. But it does not change the fundamental limitations of inquiry as audit evidence.

## 2.4 Reperformance

Reperformance is the independent execution of procedures or controls that were originally performed by the client, and is widely considered one of the most reliable forms of audit evidence. In tests of controls (ToC), it involves the auditor recreating a control activity to verify that it was done correctly and consistently. ISA 500 lists reperformance as a core audit technique, and it is particularly effective because it reduces reliance on client representations and provides direct evidence of control operation.

### AI's Role in Reperformance

AI, robotic process automation (RPA), and rule-based scripting can enhance the efficiency of reperformance — particularly for controls that leave a structured, digital trail. For example, where a control involves the matching of purchase orders, invoices, and goods received notes, AI can rerun the logic across a dataset to verify whether the same result would be

obtained. Similarly, automated recalculation tools can confirm whether reconciliations or allocations were performed accurately, according to defined rules.

**Where AI Is Less Suitable for Reperformance in ToC**

Many of the limitations seen with AI in inspection apply to reperformance as well — especially when comparing the ToC phase with risk assessment procedures (RAP): AI excels during RAP, where it can be used to analyse 100% of the general ledger, perform full-population recalculations, or replicate client processes across massive datasets to identify anomalies or control design issues. In this phase, the aim is to surface red flags, not confirm individual control operation. In ToC, however, the auditor typically works with a smaller, predefined sample — often 20 to 60 items per control. Reperformance in this context usually involves straightforward, deterministic steps: recomputing an amount, checking an automated system flag, or verifying that a control step (like review and sign-off) triggered the expected system result.

AI offers less of an advantage here for three key reasons:

1. Low Volume, Low Complexity: The sample size is small, and the steps involved in reperformance are often basic (e.g., does amount A minus amount B equal C?). The overhead of setting up AI workflows — especially if RPA is used — may not justify the time savings.
2. Limited Access to Source Documents: While AI can access full-population data from structured sources (GL, logs, control metadata), most reperformance in ToC requires comparing those system results against source documentation (invoices, approvals, bank statements). In most engagements, auditors do not have full digitized access to this documentation, and must still rely on sampling and manual inspection. AI cannot reperform what it cannot see.
3. High Setup Cost for Narrow Use Cases: AI excels at scale — but in ToC reperformance, each control type might require a bespoke setup. If Control A requires checking discount caps, and Control B requires

recalculating currency conversions, each may require separate AI logic, making it inefficient unless volumes are high.

**Where AI Still Adds Value**

There are contexts where AI and automation do help:

- Structured Digital Workflows: In fully digital environments (e.g., ERP systems with automated approval chains), AI can trace whether the expected control logic executed properly. For example, did the system block a purchase over a threshold without proper authorization?
- Pre-screening Sample Items: Even when sampling, AI can triage which items are more likely to fail based on metadata (e.g., close to cutoff, near thresholds, involving high-risk suppliers), improving sampling strategy and audit efficiency.
- Log File Reperformance: For system-based controls (e.g., access revocation, automated three-way match), AI can reperform steps by analysing system logs — assuming full access is granted.
- Back-end Recalculations: AI can assist in recomputing interest accruals, amortization schedules, or depreciation charges — especially in complex calculations that are consistent across items.

In short, while AI-powered reperformance tools are promising — particularly in RAP and in highly digital clients — their usefulness in ToC is limited by real-world constraints: the need for source documentation, low sample sizes, and task simplicity. Like with inspection, AI should be seen as a tool to support, not replace, auditor judgment and execution. Its biggest value is in automating high-volume, rules-based recalculations in structured systems — not in running deterministic tasks on small samples of semi-manual processes.

Where the client's systems are mature, and the audit firm has strong data access protocols and prebuilt AI modules, the benefits increase. Otherwise, traditional manual reperformance remains the more efficient and practical approach in most engagements.

# 3. Substantive Procedures

Substantive procedures are audit tests designed to directly verify financial statement amounts and disclosures, typically through a combination of analytical procedures and tests of details. While preliminary analytical procedures (as discussed under ISA 315) are focused on high-level scanning of relationships to inform risk assessments during the planning phase, substantive analytical procedures aim to obtain direct audit evidence — and are therefore subject to the requirements of ISA 520.

The distinction is crucial: AI's role in planning-phase analytics is about identifying where to focus the audit, whereas in substantive procedures, it helps determine whether a material misstatement exists. This shifts the bar in terms of evidence quality. Substantive analytical procedures must be more precise, predictive, and anchored in strong expectations to serve as persuasive audit evidence. In this context, AI has a powerful — but bounded — role.

## 3.1 Analytical Procedures

Substantive analytical procedures under ISA 520 require the auditor to develop expectations and investigate significant differences. AI enhances this process by:

- Generating dynamic, multi-variable expectations: Rather than relying on prior-period balances or budgets alone, AI models integrate historical trends, industry benchmarks, operational metrics, and even macroeconomic indicators to build richer expectations.
- Detecting subtle anomalies: AI can model seasonality, lag effects, and non-linear patterns to identify deviations a human might overlook.
- Improving precision: AI enhances the plausibility of expectations, allowing the procedure to stand alone or supplement tests of detail.

Example: An AI model evaluating revenue identifies an unusual pattern: Q4 sales increased 18% year-over-year, which on its own might seem positive. However, the model also factors in the following:

- Customer web traffic dropped 12% during the same period;
- Average shipping lead times, derived from logistics system metadata, increased by 40%, suggesting operational bottlenecks;
- Inventory turnover ratios, when benchmarked against industry peers, sharply declined;
- Customer service logs showed a spike in refund and complaint tickets for post-quarter shipments;
- A new deferred revenue recognition policy was implemented but inconsistently applied across subsidiaries.

The AI correlates these factors and flags the revenue increase as anomalous — suggesting possible early revenue recognition (goods shipped but not yet received), inconsistent policy application, or booking of sales unsupported by operational activity. This prompts the auditor to investigate cut-off procedures, contract terms, and potential override of controls at year-end.

To be used as substantive evidence, the AI-generated expectation must meet ISA 520's quality criteria: it must be developed using reliable data, be sufficiently precise, and include a documented threshold for what constitutes a significant deviation. The auditor must then investigate any variance, corroborate AI findings, and consider whether the result constitutes sufficient appropriate evidence.

Firms such as Deloitte (BEAT) and PwC (Halo) already use predictive models that generate expected balances for key accounts, allowing for exception-based investigation. The goal is not to replace auditor judgment, but to give them a more robust starting point from which to probe deeper.

In summary, AI enables a more analytical and insight-driven form of substantive testing. It moves the auditor beyond static comparisons and into dynamic, model-informed reasoning — provided the auditor

understands the underlying assumptions and investigates AI-flagged anomalies with the same professional skepticism as any traditional variance.

## 3.2 Risk Scoring and Outlier Identification

While AI-driven risk scoring and anomaly detection were already discussed in the context of audit planning (see Sections 1.3 and 1.4), that earlier use was diagnostic in nature — focused on identifying risky accounts, time periods, or business processes to shape the overall audit strategy. The emphasis was on understanding the entity's environment and surfacing inherent risks of material misstatement, as required under ISA 315.

In contrast, this section focuses on how similar AI tools are used at a later stage, during substantive testing of details, to select specific transactions or balances for examination. Here, the goal is no longer to determine where risk lies, but to use AI to intelligently prioritize which individual items should be tested — replacing traditional random or monetary unit sampling with data-driven, risk-weighted selection. This approach aligns with ISA 500 and ISA 530, enhancing both the efficiency and effectiveness of substantive procedures by ensuring audit effort is focused on the most anomalous or potentially misstated entries.

### Traditional Sampling and Its Limitations

Traditional substantive testing often involves random or stratified sampling, where items are selected based on monetary value, time period, or location. While conceptually sound, these methods carry an inherent risk: rare but high-risk anomalies may go undetected if not captured in the sample. For instance, a single fictitious invoice or override posted at quarter-end may be statistically invisible under standard sampling, even though it poses substantial audit risk.

### How AI Replaces or Enhances Sampling

AI offers a solution through transaction-level scoring. These tools assess each item in a dataset — such as journal entries, invoices, or payroll transactions — and assign a risk score based on multiple dimensions:

- Amount (e.g., high-dollar value transactions)
- Timing (e.g., posted near cut-off dates)
- Frequency (e.g., unusual transaction volumes by a single user)
- Nature of account pairings (e.g., debits to revenue and credits to cash)
- Override history (e.g., items bypassing standard controls)

The resulting scores allow auditors to construct a "smart sample" — not based on chance or thresholds, but on statistical inference and learned risk patterns. This reduces audit blind spots and increases the likelihood that material misstatements, including fraudulent ones, are detected.

**Techniques: Supervised and Unsupervised Learning**

Many tools rely on unsupervised learning (e.g., clustering algorithms) to determine what “normal” looks like in the client’s data. Transactions that do not fit these clusters are flagged as anomalies. Others employ supervised learning models trained on past audit findings or control failures to predict which transactions are likely problematic.

For example:

- MindBridge AI Auditor scores journal entries using dozens of risk indicators and presents a heatmap for auditors to investigate.
- PwC’s GL.ai tool uses a combination of AI models to flag transactions that resemble past fraudulent activity or show unusual behavioural patterns.

**Practical Application**

These tools do not replace testing. Rather, they direct the auditor’s attention to the riskiest items — transforming audit sampling into a targeted

investigative exercise. Instead of choosing 30 items randomly, the auditor might select:

- The 15 most anomalous items (based on AI scores)
- 10 items stratified across key accounts
- 5 items of known interest to regulators or external stakeholders

This hybrid model retains statistical defensibility while improving audit depth.

**Cautions and Limitations**

- False Positives: Not all flagged transactions are problematic. A high-risk score may simply reflect an unusual but legitimate business event.
- False Negatives: AI may overlook cleverly disguised fraud, especially if it mimics normal behaviour
- Model Transparency: Auditors must understand how the model works, what data it relies on, and what risk features it emphasizes. Black-box models that cannot be explained or justified are problematic under ISA 230 (documentation) and ISA 500 (appropriateness of audit evidence).

In summary, AI-based risk scoring and anomaly detection provide a more precise, data-driven approach to substantive testing. By helping auditors select the most meaningful items for investigation, these tools move auditing beyond blind sampling — enabling smarter, risk-focused evidence gathering that aligns with modern audit objectives and professional skepticism.

## 3.3 Automated Document Matching and Verification

Substantive procedures often require auditors to verify recorded transactions by inspecting supporting documents — a process traditionally referred to as “vouching.” For example, to test the occurrence of expenses or accuracy of revenue, an auditor might select a sample of general ledger

entries and match them to external evidence such as invoices, contracts, or delivery receipts. This work, while foundational, is often tedious and time-consuming.

**Traditional Approach: Manual Vouching**

The classic method of testing for example the occurrence or accuracy of transactions involves the auditor manually:

- Selecting a sample of transactions from the general ledger.
- Requesting corresponding source documents (e.g., invoice, GRN, bank proof).
- Comparing fields like date, amount, vendor/customer name, and reference numbers.
- Ticking that these match, and flagging any discrepancies for follow-up.

While effective, this process is constrained by the availability and quality of documentation, and the human auditor's ability to process only a limited number of samples.

**AI-Enhanced Matching: OCR, NLP, and Structured Comparison**

Where clients maintain well-organized digital documentation, AI can assist by:

- Optical Character Recognition (OCR): Converting scanned PDFs into machine-readable text.
- Natural Language Processing (NLP): Identifying key terms, dates, amounts, and counterparties within unstructured documents.
- Matching Algorithms: Comparing extracted fields with accounting records to validate that transactions are supported by adequate documentation.

For example, an AI system might ingest 2,000 invoices, extract the invoice number, amount, date, and vendor name, and attempt to match these to

purchase entries in the general ledger. Mismatches — whether in amount, missing documents, or date misalignments — would be flagged for auditor review.

This automation expands the potential scope of testing:

- Instead of manually inspecting 25 invoices, the auditor could review the exception report from a full-population test.
- This reduces the risk of missing anomalies due to sampling limitations.
- It also allows for detection of subtle irregularities, such as consistently late approvals or supplier inconsistencies.

**Use Cases Where AI Is Effective**

AI-powered document matching is particularly promising in scenarios with:

- High-volume, repeatable processes (e.g., accounts payable, payroll, contract billing).
- Structured formats, such as standardized invoice templates, PO systems, or automated delivery logs.
- Digitally native documentation, like PDFs with embedded text layers or system-generated confirmations.

**Limitations and Practical Constraints**

Despite the potential, several real-world constraints limit AI's universal applicability in this domain:

- Poor Document Quality: OCR struggles with low-resolution scans, handwritten annotations, or non-standard layouts. This creates noise, requiring manual cleanup or verification.
- Inconsistent Formats: SMEs often use free-form invoice templates or receive documents from diverse suppliers in unpredictable formats, limiting how well NLP models can generalize.
- Limited Digitization: Many audit engagements still involve physical documents or scanned PDFs without searchable text, especially in

emerging markets or smaller entities. AI-based matching cannot begin without digital inputs.

- Setup Overhead vs. Task Simplicity: In ToD, the sample sizes are often small (e.g., 30–60 items per class of transactions). Setting up AI to handle document parsing and matching may take longer than just doing it manually — especially when the check is a simple field comparison.

**AI and Fraud Detection Potential**

Although rare in basic tick-and-match procedures, AI can sometimes reveal deeper anomalies:

- Term mismatches: An invoice might show "net 60," while the accounting record suggests "net 30" — a potential sign of override or fraud.
- Document tampering: Computer vision can detect inconsistencies in fonts, shadows, or image compression — possible red flags of document forgery.
- Metadata inconsistencies: AI can detect if a document's "creation date" is suspiciously later than the transaction date — suggesting it may have been fabricated post hoc.

However, such applications are typically reserved for forensic-style engagements, not standard financial audits.

In Summary, AI has expanded what's technically possible in document verification, enabling auditors to scan, extract, and match thousands of documents against accounting records. But this potential is constrained by the fact that many clients do not yet operate fully digital environments, and the audit task itself — matching amounts and dates — is often too narrow to justify AI's overhead unless working at scale.

## 3.4 Inventory Observations and Asset Verifications

AI technologies such as drones, image recognition, and satellite imagery — previously discussed in Section 2.2 under observation — also play a role in substantive procedures when the objective shifts from assessing control performance to obtaining direct evidence of existence and, in some cases, the condition of physical assets. For example, drone footage enhanced with AI object recognition can be used to verify inventory stockpiles in agriculture or manufacturing environments. Similarly, satellite imagery may provide evidence of the continued operation or abandonment of remote facilities. However, these tools are typically supplementary. They may confirm presence or approximate quantity, but cannot assess internal condition, precise location, or obsolescence — factors often material to valuation.

Finally, whether such footage qualifies as observation or inspection depends on timing and involvement (see Section 2.2), but in a substantive context, it must support the existence assertion and meet ISA 500's criteria for sufficient appropriate audit evidence.

## 3.5 Recalculation and Compliance Checks

Traditionally, auditors perform recalculation procedures to verify the mechanical accuracy of figures computed by the client. This includes depreciation schedules, interest calculations, tax computations, lease liability models, and more. These are typically low-judgment procedures where the auditor is independently verifying a mathematical outcome based on known inputs and assumptions.

With modern software, many of these recalculations can be automated — enabling the auditor to reperform dozens of models or spreadsheets with minimal manual input. However, a key question arises:

**Is This Truly AI?**

While many recalculation tools are marketed as AI-powered, the core logic behind these procedures is usually deterministic and rules-based — aligning more closely with traditional automation than with genuine artificial intelligence. Consider the following:

- Recalculating interest expense based on a fixed rate and loan balance is basic arithmetic.
- Checking the application of depreciation schedules or tax rates follows explicit formulas.
- Replicating a lease liability model using IFRS 16 assumptions is a matter of structured logic.

These tasks are typically performed using rule-based engines or scripted automation, more akin to Robotic Process Automation (RPA) or Excel macros than machine learning. The software applies predefined logic, not adaptive reasoning.

**When AI Does Enhance Recalculation**

Still, there are edge cases where AI plays a more central role:

- Cross-checking calculated fields with external benchmarks to detect overoptimistic assumptions or misapplied models (e.g., impairment losses, bonus accruals).
- Simulating alternate scenarios to assess model robustness — for instance, re-running a DCF model with different inflation assumptions and comparing results.
- Auditing complex spreadsheets by interpreting formula logic across hundreds of linked sheets to identify contradictions or misflows.

Even in these situations, AI supports rather than replaces traditional recalculation. Its role is to enhance scope, guide focus, or surface potential errors — not to generate audit conclusions independently.

**Compliance Checking and Narrative Analysis**

The second major category under this heading is disclosure review. Here, AI's role is clearer and more substantial.

AI tools — especially those using Natural Language Processing (NLP) — can:

- Check whether all required disclosures under IFRS or local GAAP are present by matching financial statement text against disclosure checklists.
- Flag missing items (e.g., no mention of significant accounting judgments under IAS 1, or absent sensitivity analyses under IFRS 7).
- Compare current disclosures to prior years or peer companies to identify regressions, inconsistencies, or boilerplate reuse.

For example, an NLP model can review the notes to the financial statements and alert the auditor that although IFRS 15 applies to the entity, no reference to performance obligations is included — prompting a deeper review.

AI is especially effective when disclosure rules are lengthy and dynamic — as in the case of climate-related disclosures, IFRS 9 hedge accounting, or new IFRS amendments.

However, these tools are not immune to limitations:

- Disclosure logic is complex, and some required information may be embedded in other sections.
- AI can miss nuance, such as a disclosure that is technically present but inadequately explained.
- False positives may occur when AI doesn't understand firm-specific language or context.

Thus, while AI can surface potential gaps, it remains the auditor's responsibility to evaluate whether the financial statements meet the substance and spirit of the reporting standards.

This section forms a bridge to the next — where AI moves beyond recalculation into judgmental areas like accounting estimates. There, the promise — and complexity — of AI is even greater.

## 3.6 AI in Evaluating Accounting Estimates

Accounting estimates represent some of the most judgment-heavy areas in financial reporting. From expected credit losses to impairment testing and

warranty provisions, auditors are tasked not with re-performing a mathematical calculation, but with evaluating whether management's assumptions are reasonable, supportable, and free from bias. AI is emerging as a powerful tool in this space, not because it automates judgment, but because it provides auditors with a rigorous, data-driven benchmark to challenge those judgments.

**Why Estimates Are Challenging to Audit**

Accounting estimates are inherently forward-looking. They rely on uncertain inputs, probability-weighted scenarios, and subjective judgments. ISA 540 (Revised) requires auditors to assess whether management has applied appropriate methods, used reasonable assumptions, and incorporated relevant data. Auditors must consider not just accuracy, but reasonableness. For example with Expected Credit Losses (IFRS 9): Probabilities of default (PD), loss given default (LGD), and macroeconomic overlays. Or Fair Value Measurements (IFRS 13): Especially Level 3 inputs where market data is limited. These estimates often involve models developed by management or specialists. The auditor's responsibility is not to create a better model, but to evaluate whether the model is reasonable in the circumstances.

**How AI Assists with Estimate Evaluation**

AI enhances audit work over estimates by providing a "quantitative challenger" to management's models. It does not replace auditor judgment or override management estimates, but offers a second lens grounded in broader datasets and pattern recognition.

Here are some examples where AI can contribute:

*Expected Credit Losses (ECL)*

- Analyse historical payment patterns, borrower profiles, and macroeconomic variables to estimate PD and LGD.
- Model multiple forward-looking scenarios (base, adverse, upside) as required by IFRS 9.

- Flag segments of the loan portfolio where the allowance may appear too lenient or too conservative relative to economic conditions.

*Net Realizable Value (NRV)*

- Forecast future selling prices by analysing recent sales velocity, inventory aging, discount trends, and external pricing data.
- Identify items at risk of overvaluation due to obsolescence or seasonality.

*Impairment Testing (IAS 36)*

- Assisting in DCF model inputs, such as revenue growth, margins, and discount rates based on market data.
- Benchmarking management's assumptions (e.g., terminal growth vs GDP, discount rates vs WACC).
- Highlighting overly optimistic assumptions that deviate from industry or historical norms.

*Provisions and Legal Reserves (IAS 37)*

- Analyse past claim patterns and legal case outcomes.
- Scan legal correspondence and litigation databases to identify risks not included in provisions.

*Fair Value (IFRS 13)*

- Identify comparable transactions in market databases.
- Build alternative valuation models based on industry and economic inputs.
- Flag assumptions that fall outside peer or market ranges.

In each case, the AI does not conclude whether an estimate is acceptable, but presents an alternative view that can trigger auditor follow-up.

**Auditor's Role: Interpretation and Professional Skepticism**

AI-generated outputs are not audit conclusions. The auditor must:

- Assess whether the AI's methodology and assumptions are appropriate.
- Compare outputs to management's assumptions and investigate discrepancies.
- Document how the AI output influenced their evaluation, consistent with ISA 540.
- Apply skepticism, especially in areas of high estimation uncertainty or management bias.

A helpful analogy is that AI acts like a skeptical team member offering a second opinion. The auditor must weigh that opinion, not blindly follow it.

**Black Box Models and the Need for Explainability**

Auditors cannot rely on opaque AI models. Under ISA 230 (Audit Documentation) and ISA 500 (Audit Evidence), conclusions must be supported by explainable logic. This presents a challenge:

*Where AI Is a Black Box*

- LLMs like GPT: May generate reasonable-sounding text, but decisions are based on token prediction, not structured analysis.
- Neural networks in fraud or risk scoring: May identify risks without a clear explanation of which variables triggered the flag.
- Computer vision models: May estimate stock levels from images without explaining the methodology.

If a model cannot explain its decision path in human terms, it cannot be relied upon as audit evidence.

*Where AI Is Transparent and Auditable*

- Decision Trees / Random Forests: Show which variables matter and how decisions are made.
- Rule-based NLP systems: Clearly extract information from structured text.

- Explainable AI (XAI) Tools: Layers like SHAP and LIME explain model outputs by showing variable contributions.

The auditor should choose AI tools that offer transparency, understand how inputs become outputs, and document the rationale behind AI-informed conclusions.

**Could Empirical Reliability Replace Explainability?**

Today, black-box reliance is restricted. But could the future allow for exceptions if the models show consistent empirical accuracy?

If a black-box model (say, a deep neural net that forecasts expected credit losses or flags potential frauds) consistently outperforms human auditors or traditional models over years, and if regulators begin to view track record as evidence, then that might lay the groundwork for acceptability.

Imagine a world where:

- Black-box Model X has been validated on 10 million real-world accounting scenarios,
- Its predictions were later shown to be more accurate than human estimates 98% of the time,
- It is rigorously monitored by a third-party regulatory framework,
- It comes with model drift detection and bias mitigation systems.

Under such conditions, the model's empirical trustworthiness might begin to rival, or even outweigh, the need for explainability.

In medicine, we already rely on "black-box" systems in diagnostic imaging or drug discovery — not because we fully understand how they work, but because they work better than alternatives. This might set a precedent.

If the profession shifts toward risk-weighted evidence models, auditors might one day say:

"The AI assigns an 87% probability that the impairment estimate is understated — and its long-term accuracy rate exceeds 90% under similar

conditions. Based on this, we determined additional evidence was necessary."

That's not a deterministic explanation — but it could be acceptable in a risk-based audit evidence framework, especially if backed by a robust audit trail of the AI model's development and validation.

**Counterpoints: Why the Current Standards Still Matter**

*Accountability Requires Traceability*

If an audit conclusion is ever challenged (e.g., in litigation), saying "the AI said so" is not defensible — not today, and probably not for decades. Human auditors are the accountable party, and you cannot cross-examine a neural net.

*Bias and Drift Risks*

Even highly accurate models can make systematic errors if the underlying data is biased or if the model begins to drift due to new market dynamics. Without explainability, you might not detect these errors until they cause harm.

*Audit Evidence Is a Legal Construct*

Audit evidence isn't just a technical standard — it's embedded in legal and ethical frameworks. The bar for "appropriateness" is not only accuracy but also understandability by third parties, including courts and regulators.

Instead of abandoning explainability, the future is more likely to involve:

- Explainability Layers (like SHAP or LIME) built over black-box cores, giving enough insight to pass ISA requirements.
- Model Assurance becoming a field of its own — where models are audited, version-controlled, and assigned "trust scores."
- Human–AI collaboration standards, where AI can inform, but not replace, professional judgment — unless validated by third-party review bodies.

# 4. Audit Reporting and Completion

The final phase of the audit process involves forming conclusions, drafting the auditor's report, and communicating findings (such as management letters or internal control deficiencies). While this phase is more qualitative and judgment-driven, AI can still contribute to improving accuracy, efficiency, and thoroughness in several ways.

**AI-Assisted Drafting and Review of Audit Reports**

Modern AI language models (like GPT-based systems) can assist in drafting standard sections of audit deliverables and communications. For instance, AI tools can generate first drafts of management letters or audit committee reports based on structured input from the audit file. These tools save time on boilerplate language and improve consistency in tone and format across engagements. However, caution is critical: audit reports and auditor communications are governed by precise professional standards. Auditors must thoroughly review any AI-generated text, ensuring that terminology, tone, and conclusions align with ISA requirements. At present, most firms limit AI-generated drafting to internal memos, working paper summaries, or documentation aids — with the final editing and approval resting squarely with the human auditor.

**Automated Consistency Checks Across Financial Statements**

AI systems can significantly streamline the traditional "tick-and-tie" process by automatically scanning the financial statements, notes, and accompanying reports to verify numerical and textual consistency. For example, if lease obligations are disclosed in Note 12, the AI will confirm the same figures appear in the balance sheet and that all cross-referenced tables (like maturity analyses) match. It can also identify inconsistencies in prior-year restatements, incorrect rounding, or discrepancies between ratio disclosures and actual calculations. This not only reduces clerical effort but

helps catch errors that could undermine the credibility of the report. These consistency checks replicate the meticulous diligence traditionally handled by experienced staff — now enhanced by real-time AI cross-verification.

**NLP-Based Compliance and Disclosure Reviews**

Auditors must ensure that the financial statements comply with all relevant accounting and regulatory requirements. AI, particularly NLP (Natural Language Processing) models, can maintain an up-to-date checklist of required disclosures (e.g., IFRS, GAAP, jurisdictional standards) and scan financial statements to confirm whether each requirement is addressed. For example, if IFRS 15 mandates a description of how performance obligations are identified, the AI can search for relevant disclosures and flag their absence. These tools can also compare disclosures against prior years or peer companies, highlighting missing policies or boilerplate language. While human oversight remains essential for judgment-heavy narratives, AI ensures completeness and coverage of technical requirements across lengthy disclosure documents.

**AI in Quality Control and Engagement Reviews**

During the engagement completion phase, many firms require an Engagement Quality Control Review (EQCR) by a senior reviewer. AI can assist this process by compiling summaries of key findings, highlighting areas with elevated risk or significant audit adjustments, and flagging incomplete sign-offs. For example, an AI dashboard might list all override decisions made by the audit team (e.g., where an auditor rejected an AI-suggested sample), all material estimates audited, or sections where assumptions were heavily challenged. This streamlines the reviewer's task by focusing attention on judgment-heavy areas. AI can also cross-check that all required procedures for high-risk accounts were performed, and that documentation meets firm and ISA standards. While AI does not replace professional oversight, it acts as a quality enhancement tool to ensure audit completeness and focus.

**AI Learning Loops: Improving Future Audit Cycles**

While not directly tied to issuing the current year's audit report, AI has a unique capacity to improve future audits through continuous learning. Audit firms can feed anonymized data from completed engagements — including known misstatements, control deficiencies, or client risk factors — into machine learning models to refine their performance. For example, if this year's audit uncovered premature revenue recognition related to performance obligations, future AI models can be trained to flag similar revenue arrangements in the planning phase of other clients. Similarly, labeling certain anomalies as errors (or not) helps fine-tune AI anomaly detection tools. In this way, each audit cycle contributes to a virtuous learning loop, where firmwide audit quality improves through cumulative insights.

Overall, the audit reporting and completion phase remains grounded in auditor judgment and professional standards. Yet AI enhances this phase by reducing clerical burden, catching inconsistencies, and supporting reviewers with summarized insights. Importantly, the auditor — not the AI — retains responsibility for forming the audit opinion. The power of AI lies in its ability to augment human diligence, not to replace it. With appropriate safeguards, AI can help ensure that the final product of the audit — the auditor's report — is as accurate, complete, and professional as possible.

---

# 5. Implications of the Client's Use of AI

Artificial intelligence is not only transforming the auditor's toolkit — it is increasingly embedded within the client's own financial processes, systems, and controls. From machine-learning credit models to automated invoice matching systems, clients are adopting AI in ways that impact transaction processing, internal control, and financial reporting. This poses new challenges — and opportunities — for the auditor.

ISA 315 (Revised 2019), ISA 330, and ISA 500 require the auditor to understand and evaluate the design and implementation of the client's internal controls

and assess the reliability of information produced by the entity. When AI systems are part of those controls or data sources, auditors must treat them as integral to the audit risk model.

## 5.1 Understanding AI-Driven Processes and Controls

As part of risk assessment procedures (RAP) under ISA 315 (Revised 2019), auditors are required to obtain an understanding of the client's internal controls relevant to the audit. When clients use AI-based systems in processing transactions or managing controls, the auditor must:

- Understand the design of the AI control: What does it do? Which financial assertions does it support (e.g., completeness, accuracy, cutoff)?
- Understand its intended role in the control environment: Is it preventive or detective? Manual or automated?
- Evaluate whether it is implemented: Is the system live? Is it integrated with existing systems or run offline? Is management using its outputs in decision-making?

The goal is not to test whether the AI works, but to understand whether it could work — i.e., that it is designed to address risks and is functioning in concept.

Examples include:

- A neural network that flags unusual purchase orders for manual review;
- An AI-based contract reader that extracts lease terms for IFRS 16 calculations;
- An automated credit model that sets provisioning rates based on borrower profiles.

Here, the auditor asks:

- How does the AI make decisions? Is it rule-based, trained on past data, or adaptive?

- What are its data sources? Internal only? External feeds?
- Who reviews or overrides its outputs? Is there a documented process for human oversight?
- Are safeguards in place to prevent erroneous data or model drift?

Understanding these factors helps auditors determine whether the AI system introduces new risks (e.g., automation bias, model error, lack of audit trail), and whether it can be considered a control relevant to the audit under ISA 315. If so, the auditor may choose to test the operating effectiveness of this AI control — but that is a separate, later step under ISA 330.

## 5.2 Evaluating Effectiveness of AI Controls

If, after risk assessment, the auditor decides to rely on an AI-based control, they must go beyond understanding its design — they must test whether it actually operates effectively throughout the audit period.

Testing an AI control's operating effectiveness could involve:

- Submitting dummy transactions (if feasible) to see how the AI responds;
- Reviewing exception reports to confirm that flagged items were handled correctly;
- Verifying retraining processes or model updates during the period (if the system evolves);
- Assessing user access and change logs to ensure the AI system wasn't tampered with.

The question is no longer "Is the AI well-designed?" but "Does it consistently function to prevent or detect misstatements?"

If the control fails, or cannot be tested due to its opacity or lack of logging, the auditor must adjust the audit strategy — typically by increasing substantive testing or using alternative procedures.

Many AI controls may be poorly documented or lack traditional audit trails. In such cases, ISA 500 (reliability of evidence) and ISA 230 (audit documentation) become critical. Without sufficient audit evidence of effectiveness, reliance is not justified — no matter how sophisticated the AI system may be.

## 5.3 Risk of Automation Bias

Automation bias — the over-reliance on AI outputs without sufficient human skepticism — can infect both the client and the auditor. ISA 240 and ISA 500 caution against uncritical acceptance of management assertions or system-generated information.

At the client level, this may manifest as: "We trust the AI's estimate — it always gets it right." At the auditor level, it could be: "Their AI ECL model is advanced, we assume it's fine." Both are unacceptable under auditing standards.

Auditors should:

- Inquire how management reviews AI outputs and escalates anomalies,
- Review any post-hoc validation or back-testing,
- Document how they challenged the AI's assumptions and inputs,
- Avoid sole reliance on AI tools as substantive evidence without traditional corroboration.

Even if the AI system performs well, the auditor's role is not to accept, but to interrogate.

## 5.4 Audit Evidence and AI-Generated Information

ISA 500 requires the auditor to evaluate the reliability of audit evidence. If a client uses an AI model to generate key estimates (e.g., ECL, inventory obsolescence, DCFs), then that AI model itself becomes part of the evidence chain — and must be scrutinized like any expert or tool.

Auditors should consider:

- The appropriateness of the AI's assumptions,
- Completeness and accuracy of its inputs,
- Transparency of the decision-making process,
- Comparability to external evidence (benchmarks, actual outcomes),
- Whether the system retains an audit trail (e.g., what parameters were in place at year-end?).

Where the AI is too opaque, or lacks adequate validation, the auditor may default to their own independent estimate and treat the AI as unsupported.

## 5.5 Changes to Audit Approach

The presence of AI may raise or reduce assessed risk, depending on how it is implemented. Key planning decisions impacted include:

- Whether the control environment is stronger or weaker due to AI,
- Whether the AI has a proven track record or is newly deployed,
- Whether the AI logs its decision path and audit trail,
- Whether model drift or frequent retraining compromises consistency during the audit window.

ISA 330 and ISA 520 may require modification of testing strategies — e.g., increasing substantive testing where the AI control is untested or lacks transparency, or reducing testing where the AI consistently detects errors the client addresses. But such adjustments must be justified by sufficient understanding and testing — not by convenience or superficial appearances.

The overarching principle: AI is not immune from audit scrutiny — if anything, it invites deeper skepticism. Unless the AI is explainable, validated, and clearly documented, it cannot be blindly trusted or used as a substitute for human judgment. But if properly assessed, AI-enabled systems may

enhance both the client's reporting reliability and the auditor's assurance strategy — enabling more efficient, focused, and intelligent audits.

---

# 6. Professional Scepticism

Professional skepticism is a cornerstone of the audit process. According to ISA 200, it is "an attitude that includes a questioning mind, being alert to conditions which may indicate possible misstatement due to error or fraud, and a critical assessment of audit evidence." This foundational mindset underpins every phase of the audit and ensures that auditors maintain independence of thought, avoid undue reliance on client representations, and seek persuasive evidence before forming conclusions.

Critically, professional skepticism is not just about suspecting fraud — it's about maintaining an objective distance and never assuming that things are correct simply because they appear to be or because the client has a history of honesty. ISA 240, which focuses on fraud, warns against allowing familiarity with management to reduce the auditor's critical stance, stating that a skeptical attitude is essential regardless of perceived integrity.

Beyond the standards, a robust academic body has explored professional skepticism, offering a more granular understanding of its components. A landmark contribution is Hurtt (2010), who introduced the Hurtt Professional Skepticism Scale (HPSS). This scale identifies six key traits:

1. Questioning mind – a curiosity-driven tendency to ask probing questions rather than accept information at face value.
2. Suspension of judgment – the capacity to delay conclusions until adequate evidence is obtained, avoiding premature closure.
3. Search for knowledge – a drive to investigate beyond superficial evidence, seeking corroboration or clarification.

4. Interpersonal understanding – an awareness that people have motives and biases, leading to skepticism even in the face of likeability or charisma.
5. Autonomy – independence in thought and resistance to social pressure or authority, allowing the auditor to hold contrarian views when needed.
6. Self-confidence – the assurance to challenge explanations, pursue anomalies, and remain firm when evidence demands it.

These traits combine cognitive, social, and emotional capabilities. In practice, they guide auditors to investigate more thoroughly, weigh evidence critically, and avoid overreliance on client-provided information.

Other researchers, such as Nolder and Kadous (2018), have suggested a split between skeptical mindset (the analytical component) and skeptical attitude (the affective and motivational component). The mindset involves cognitive tools such as generating alternative hypotheses and evaluating evidence from multiple perspectives, while the attitude reflects the degree to which the auditor is concerned about risk and is motivated to follow through on that concern.

**Where AI Aligns with Professional Skepticism**

AI systems show considerable potential to support the application of professional skepticism — particularly in its cognitive and procedural aspects:

- Critical analysis and anomaly detection: AI excels at processing large data sets and identifying unusual patterns or inconsistencies that a human might miss. For example, AI can compare invoice data across an entire population to flag duplicates, unmatched amounts, or temporal inconsistencies — directly supporting the requirement to perform a "critical assessment of audit evidence."
- Scalable pattern recognition: Machine learning models trained on historical fraud or misstatement cases can flag subtler, systemic

anomalies — functioning as a kind of institutional memory that mimics the pattern-recognition intuition of experienced auditors.

- Bias resistance: Properly designed AI systems do not suffer from common human cognitive biases (e.g., confirmation bias, anchoring, or overconfidence). This could allow AI to act as a consistency-checking layer — treating every transaction impartially and applying the same verification rigor, regardless of who submitted the data or the auditor's familiarity with the client.
- Procedural replication of skeptical behaviour: Audit firms already attempt to enforce skeptical behavior procedurally (e.g., mandatory thresholds for additional testing). Similarly, AI can be programmed to follow such logic rules rigidly and consistently, ensuring skeptical follow-ups are not skipped due to fatigue, pressure, or personal bias.
- Knowledge access and hypothesis generation: Large language models, trained on extensive professional content, can generate plausible alternative explanations for results, suggest follow-up questions, or identify inconsistencies between management explanations and prior disclosures. This reflects the cognitive component of skepticism described in the literature.

**Where AI Falls Short**

Despite these strengths, many believe that significant limitations remain in the application of AI to professional skepticism:

- Lack of a true questioning mindset: AI lacks a consciousness-driven attitude. It does not spontaneously entertain doubt or experience intuitive discomfort — it acts only when explicitly trained to respond to anomalies. It does not “wonder” whether something is wrong unless a measurable pattern prompts it.
- Absence of contextual awareness and common sense: While AI can detect anomalies, it often lacks the real-world context to evaluate their meaning. For instance, it might flag an unusual fluctuation without understanding whether it’s a genuine issue or an expected

seasonal trend. The nuanced human interpretation — shaped by experience, industry knowledge, and conversation — is difficult to replicate.

- Inability to exercise judgment: Skepticism isn't just about flagging issues — it's about prioritizing them and deciding when to escalate. AI can suggest risks but cannot determine materiality, weigh conflicting evidence in an unstructured environment, or judge when "enough" evidence has been obtained.
- Lack of social and ethical responsibility: A skeptical human auditor may dig deeper not just because of procedure, but out of a sense of duty to the public, shareholders, or regulatory standards. AI does not carry such obligations and cannot be held ethically accountable.
- Skepticism toward the AI itself: Ironically, a skeptical auditor must also be skeptical of the AI. Just as we challenge management's assumptions, we must scrutinize AI outputs: Did the model use outdated data? Was the training set biased? Are we mistaking algorithmic precision for relevance?

**Counterarguments and Rebuttals: Rethinking AI's Potential**

While the limitations above are legitimate, they deserve careful scrutiny:

*Is human intuition really superior?*

The human "feeling" that something is off is likely a product of subconscious pattern recognition. Machine learning systems, trained on vastly more data, might outperform such intuition — not in reproducing the feeling, but in surfacing the anomaly itself. What we call intuition could be modeled computationally.

*Does AI need a 'mindset'?*

ISA 200 speaks of an "attitude of skepticism," but this may be an anthropocentric framing. If we consider what the mindset accomplishes — heightened alertness and refusal to accept weak evidence — an AI system

trained to verify rigorously and flag anomalies is functionally achieving those ends. It doesn't need doubt, just precision and breadth.

*Human judgment is biased too*

Auditors are warned about client bias — but their own bias is rarely addressed with equal force. AI bias (e.g., skewed training data) is at least measurable and correctable. Human biases are subtle, opaque, and systemic. AI could thus bring a kind of objectivity that complements or even exceeds human judgment in consistency.

*Fuzzy ≠ unreplicable*

The inability to formally define every step of professional skepticism doesn't imply it's unreplicable. Many past beliefs about uniquely human capabilities (e.g., chess strategy, medical diagnostics) have been overturned by advances in AI. It is plausible that the cognitive components of skepticism — identifying red flags, asking good questions, seeking corroboration — could be increasingly automated.

*Modern AI is not rule-based*

Much criticism of AI's limitations is based on outdated rule-based paradigms. Today's AI systems use statistical inference and latent representation modeling, not IF-THEN logic. This makes them particularly suited to mimicking complex, fuzzy human behaviors — including those that underlie professional skepticism.

In conclusion, professional skepticism includes both intangible traits and replicable actions. While AI may not currently embody a skeptical "attitude" in the human sense, it can already replicate — and potentially improve upon — many of the behaviors and processes associated with skepticism. The cognitive tools of skepticism (analysis, doubt, corroboration, alternative explanations) are increasingly programmable, and in some domains, AI's speed, precision, and objectivity may exceed human performance. The affective and ethical dimensions of skepticism, while still central to traditional auditing, may not be essential for effective skeptical behavior in all contexts.

Rather than defaulting to the view that AI will merely "augment" human skepticism, it may be more productive to ask whether some aspects of judgment long seen as uniquely human are in fact more accurately, consistently, and reliably exercised by machines. The future may involve not just collaboration between human and AI auditors, but a rethinking of who — or what — is best positioned to exercise professional skepticism in its most effective form.

---

# 7. AI-Powered Audit Tools in Practice

Audit firms across the globe have been investing heavily in AI and advanced analytics tools to support their auditors. These tools often correspond to specific phases of the audit (as discussed above) and highlight the practical ways AI is already embedded in modern audit methodologies. Below we mention some prominent examples of emerging AI-driven audit tools used in practice by large firms and what they do:

## 7.1 Deloitte

Deloitte's proprietary AI tool Argus is designed to review and analyse documents and data at scale. It uses machine learning to scan contracts and financial documents, extracting key information and identifying anomalies or outliers with high accuracy. In practice, Argus might be used to review all lease agreements to find terms that deviate from standard or to analyse every journal entry for unusual descriptions. By doing so, it helps Deloitte auditors focus on the exceptions that matter. Argus essentially brings a cognitive lens to document review and transaction analysis, complementing the auditor's judgment with exhaustive data coverage.

In addition to Argus, Deloitte has developed a suite of AI tools. Cortex, for example, focuses on large dataset analysis to find patterns and irregularities, aiding in risk assessment and fraud detection. Another tool, Omnia DNAV (Digital Net Asset Value), automates the verification of

investment fund NAV calculations (important in auditing asset management clients). Deloitte also uses tools like BEAT for predictive analytics benchmarking and Optix for data visualization in audits. These tools showcase how a Big Four firm is incorporating AI at various points: from planning analytics to specialized substantive tests and visual analytics for decision-making.

## 7.2 KPMG

KPMG Clara is KPMG's integrated, cloud-based smart audit platform that incorporates data analytics and AI for auditors. Clara provides real-time collaboration and dashboards, but more importantly, it embeds AI capabilities to enhance risk assessment and analysis. For example, KPMG Clara can integrate AI routines that analyse the client's full ledger and highlight unusual transactions or perform intelligent ratio analysis comparing the client to industry data. According to KPMG, Clara's AI features provide real-time insights and risk assessments, helping auditors make better decisions during the audit. Clara can also help with documentation by automatically organizing working papers and even checking the completeness of audit procedures performed. KPMG has also partnered with technology companies (like IBM Watson in earlier years) to build AI solutions for tasks such as analysing financial contracts and correspondence for risks. The result is a platform that attempts to put AI at the auditors' fingertips throughout the audit process, rather than as a separate tool.

## 7.3 PwC

PwC has integrated AI into its audit through tools like Halo (which was an analytics suite for journal entries and other data) and more futuristically GL.ai. In collaboration with H2O.ai, PwC developed GL.ai to apply AI algorithms on clients' general ledger data. The tool can scan millions of journal entries and use machine learning to pinpoint unusual entries that warrant investigation. PwC reported that this allows their audits to coverage a much higher percentage of transactions and helps in detecting potential fraud indicators (like an employee creating and approving the same journal

entry, or entries just below authorization limits). GL.ai works alongside PwC's standard audit software (Aura) to enhance the speed and precision of audits. By automating the identification of risky transactions, it lets human auditors concentrate on evaluating those specific cases (i.e., applying their judgment to the exceptions rather than spending time doing broad data sweeps). PwC has noted improvements in audit efficiency through these tools, demonstrating how AI can transform auditing roles to be more data-driven.

## 7.4 EY

EY has similarly invested in analytics through its Helix suite, which includes modules for revenue analytics, journal entry testing, and inventory analysis, some of which incorporate AI techniques. EY also famously explored the use of drones and computer vision for inventory observations. While EY's branding for their AI in audit isn't as singularly named as Argus or Clara, they have piloted various AI applications. One example is using AI to read lease contracts for the new lease accounting standards – an exercise where EY leveraged AI to help extract relevant data from thousands of lease agreements for their clients (this was both an advisory and audit support use case). EY's global audit platform Canvas integrates these tools to allow auditors to run analytics on full populations of data. For instance, Canvas with Helix can test entire journals and flag anomalies similar to other firms' tools. EY also has an AI tool called AI Auditor (in some regions) that uses AI to analyse a client's datasets and detect anomalous transactions or entries, akin to MindBridge's approach. While each Big Four firm has its own suite, their capabilities often converge on the same idea: leverage AI to achieve comprehensive analysis and surface the items that need human attention.

## 7.5 MindBridge AI Auditor

Outside the Big Four's proprietary systems, MindBridge is a notable third-party platform making waves in auditing. It's used by many accounting firms (and was even piloted by the CPA Canada for audit innovation). MindBridge's AI Auditor applies a combination of rule-based and machine

learning algorithms to financial data to flag unusual transactions. It looks at attributes like unusual times, amounts, account combinations, and benford's law deviations, etc., and gives each transaction a risk score. Auditors can then focus on the highest risk transactions regardless of where they occur in the accounts. The benefit is a more objective way to pick samples or areas for testing – essentially, AI helps guide auditors to where the misstatements are most likely to lie. This tool embodies the future of audit evidence gathering: examining entire populations efficiently and effectively.

## 7.6 IBM's Watson and Others

KPMG and other firms have experimented with IBM Watson's cognitive capabilities, particularly for reading and interpreting documents. For example, in the audit of financial institutions, a massive number of customer contracts or loan agreements might need reviewing for terms and compliance. Watson's AI can ingest these documents and classify them, highlight risky provisions, or compare them against standard terms. In the list of top tools, IBM Watson is noted as assisting in analyzing financial documents and transactions, aiding in compliance and fraud detection efforts. While not an audit firm tool per se, this technology has been incorporated via partnerships (e.g., KPMG's alliance with IBM). Similarly, some audit teams use AI-assisted OCR tools to verify data in documents like bank statements or invoices against the accounting records automatically.

These examples illustrate that AI in auditing is not theoretical – it's already being used in practice to various extents. The tools often combine multiple technologies (machine learning, NLP, automation, visualization) under the umbrella of "AI" to tackle audit tasks that involve big data or complex analysis. A common thread is that they increase audit coverage and depth: auditors can look at more of the client's data (even 100% of transactions) and do so more intelligently. They also often integrate into existing audit workflows (like Argus and GL.ai feeding results into traditional audit documentation systems). For senior students, familiarity with these tools is valuable, as the next generation of auditors will likely work side by side with

AI-driven systems. Understanding what these tools do also demystifies AI in audit – for example, knowing that "Argus will highlight odd contract terms for me" or "Clara will visualize the risky accounts this year" makes it clear that the auditor's role is still to interpret and act on those insights.

---

# Final Reflections

Auditing has always been a profession of balance — between risk and assurance, evidence and judgment, detail and materiality. The introduction of Artificial Intelligence into this delicate ecosystem has not merely added a new tool. It has fundamentally altered the terrain.

Throughout this document, we have explored the many ways in which AI augments the audit process: enhancing risk identification, transforming sampling, enabling full-population testing, parsing unstructured documents, and even acting as a quantitative challenger to management's most sensitive estimates. But behind each of these applications lies a broader truth: AI is not simply a better calculator — it is a different kind of intelligence altogether.

It does not reason like a human, nor explain like a human, nor learn like a human. And yet, it increasingly achieves results that challenge the primacy of human judgment in areas we once thought untouchable. The idea that a machine can analyse hundreds of contracts in minutes, identify subtle risks buried in management disclosures, or project economic forecasts that rival human analysts — this would have seemed implausible to most auditors even ten years ago.

But this is only the beginning.

**AI as a Non-Human Audit Entity**

Auditing standards were never written with non-human reasoning in mind. ISA 500 tells us that audit evidence must be "sufficient and appropriate," but what happens when an AI arrives at a risk assessment we cannot fully trace? ISA 230 requires us to document the basis for conclusions — but what if that

basis is encoded in the latent space of a neural net, rather than a spreadsheet formula?

As we saw in our discussion on explainability, auditors must be able to defend their decisions. But in time, AI models may evolve to be empirically superior yet epistemologically opaque — meaning they consistently outperform traditional methods, but we cannot fully explain how they do it. Do we trust such a model? Do we reject it purely because we don't understand it? These are not theoretical questions — they are the shape of dilemmas to come.

**From Assistant to Arbiter?**

There is a real possibility that AI will move beyond being a "smart assistant" and begin to rival or even outperform the professional judgments of experienced auditors in areas like fraud detection, revenue recognition modelling, or fair value estimation. And it will not do so by replicating human logic — it will do so by leveraging pattern recognition, statistical correlation, and data richness at a scale no human brain can match.

If auditors continue to treat AI as a glorified calculator or search engine, they will misunderstand its potential — and likely misapply its outputs. This is a beast of a different nature: powerful, predictive, and increasingly autonomous in its decision-making.

That autonomy raises ethical and regulatory questions not yet answered. If an AI identifies a red flag that the human auditor overlooks, who bears responsibility? If an AI suggests a materially different impairment charge, is it negligence not to pursue it?

We do not yet have a framework for these questions — but we will need one.

**The Audit's Purpose Will Be Redefined**

There is a deeper philosophical shift underway: as AI automates more of the "what" and "how" of audit procedures, the auditor must shift focus to the "why." The role of the auditor may evolve from evidence-gatherer to evidence-curator, systems challenger, and AI overseer.

The audit of the future may be less about ticking transactions and more about validating algorithms. The evidence file may contain fewer

spreadsheets and more model explainers. The critical skills may be less about IFRS nuances and more about understanding how machine learning interacts with real-world complexity.

We are not just modernizing the audit — we are mutating it.

**Caution: AI Will Not Wait for Us**

Importantly, the pace of AI development will not wait for the auditing profession to catch up. Model capabilities — including general reasoning, synthetic data generation, and adaptive decision-making — are improving at a logarithmic rate. Audit methodologies, regulatory oversight, and educational curricula are evolving linearly, at best.

If auditors do not engage with this transformation critically and proactively, they risk being replaced rather than being augmented.

But those who do engage — who understand both the limits and leverage of AI — will shape the future of audit in their image. Not by resisting the tide, but by surfing it with integrity, judgment, and purpose.

---