



P-ISSN:3006-0265 | E-ISSN:3007-4681. Published by the Open University of Kyzylorda

Copyright © Author(s). This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC-BY\)](https://creativecommons.org/licenses/by/4.0/).
<https://bulletin.ouk.kz/index.php/bulletin/index>



AUDITING THE RELIABILITY OF GENERATIVE AI-GENERATED ACCOUNTING AND FINANCIAL DOCUMENTS: A PROPOSED CONCEPTUAL FRAMEWORK FOR EVALUATING DIGITAL AUDIT EVIDENCE UNDER ISA 500

Wisam Khalaf Nejres

University of Kirkuk, Administration & Economics

Corresponding author e-mail: wisamkhalaf@uokirkuk.edu.iq

Abstract. This work presents a framework for evaluating the accounting and financial documents generated or significantly altered by generative artificial intelligence (AI) and submitted as audit evidence. 100% human-written – the use of AI itself is not problematic. It is uncertain evidence which creates issues – specifically when it is impossible to determine the origin, input, production history or approval of a document. Through an integrative structured review, normative analysis, and conceptual synthesis, we develop the AI-Generated Document Audit Reliability (AGDAR) framework. The framework monitors the document's journey through a lifecycle in six linked dimensions (provenance, integrity, traceability, independent corroboration and accountability) consistency. The primary anchors of provenance and independent corroboration; the other dimensions connects knowledge of origin to verification of content and responsibility. AGDAR delineates four levels of AI involvement and correlates these to an incremental audit response. The paper claims that an AI-generated document might not be untrustworthy per se. Strength of the evidence is dependent upon the extent of the production chain that can be verified, and the existence of independent evidence that removes doubt. The framework enables auditors to systematically identify breaks in the chain of evidence, select proportionate procedures and reduce reliance on the appearance of documentary evidence alone.

Keywords: Generative AI; audit evidence; document reliability; content provenance; professional skepticism; ISA 500; synthetic documents.

Introduction

Generative AI is transforming the nature of accounting and financial documentation, moving certain portions of the production process away from the relatively structured, human-directed recording of content to something that can be automatically generated, revised or recomposed. Conventional accounting systems typically produce documents that are generated according to pre-established rules, templates and processing paths. As a result, the auditor will often be able to trace an output back to a recorded event and an identifiable system process. Generative systems function differently, and the outputs can differ based on the prompt, model version, retrieved information, and associated context despite an apparent similarity of task [1], [2]. Such flexibility allows them to generate standard records, descriptive narratives, interpretations, predictions and other forward-looking content [3], [4]. The problem with the audit is thus more complex than checking whether the end document looks correct. The requirement also compels the auditor to comprehend how the content was developed, what information supported its creation and whether the relationship to the underlying transaction can be verified independently [5].

A document may possess some visual or linguistic features normally associated with credible evidence, but its socio-economic or legal basis is of unknown origin due to its lack of a foundation subject to independent verification. We can expect to see vendors, invoices, contracts, signatures, graphs and financial commentary generated by systems that facilitate reliance on appearance [6], [7]. For instance, an invoice may have a fake supplier name, believable numbers and figures, but no actual purchase record, proof of delivery, or record of payment. A good contract will imply an agreement between the parties even if that agreement is not authorized or on record. A cash flow analysis can feature convincing explanations and charts that are professionally done. However, it may also rely on incorrect or false information. The document has the ability in each case to deflect the audit trail from the event and make it difficult to identify the responsible party. A professional presentation in and of itself should be considered a characteristic and therefore certainly not proof of authenticity or occurrence. The auditor should link the document to independent records, responsible parties, and the transaction it seeks to represent [8], [9], [10].

These days visual inspection or familiarity with formal features no longer satisfy conventional documentary procedures. According to Martinez [6] and Chandra et al. [7], the generative systems manage to recreate organisational logos, typefaces, signatures and institutional language as well as legal or economic writing to the extent that the document appears complete to the human eye. More importantly, the system was capable of generating a fictional narrative that contained its own fictitious context in which the names, dates, amounts, explanations and the narratives all corroborated each other. An outer check can verify if the document is of a requisite format, but not whether it is from a valid source or whether the transaction referred to has indeed occurred. An extension to the range of the investigation, including digital provenance, metadata, logs of systems and edits, and evidence of who wrote, reviewed and approved the document C2PA [9]; FRC [11] Independent verification and reconciliation between the document and financial and physical flows (for example, payments, inventory movements, delivery records, or other third-party data) should also be established [8], [5]. The aim is to transfer confidence away from what you see – based on looks for instance – to confidence based on verifiable and independently auditable chain of evidence. Processes are important when assessing the reliability of the document, however, if the auditor cannot determine who prepared the document, on what information was the document prepared and how that document was subsequently altered, does not help us to establish the content. While it may contain helpful information, it lacks an easily verifiable production process and the claims are built upon a broken digital trail. In such situations, the document is likely to become a cause of audit risks as depending on it might conceal errors, forgery or unauthorized alteration [5], [6] rather than a possible mitigating factor for such risks. It's wrong to reject something outright just because AI may have generated it. The auditor should first perform further procedures aimed at the origin of the uncertainty, in accordance with the application of professional scepticism and the evaluation of information used as audit evidence [8], [11]. External verification, checking physical or transactional evidence, checking metadata and logs, reconciliation with financial or operational flows, and checking inconsistencies between the document with one or more other documents [9], [10] can be included. The document must not be treated as persuasive audit evidence if the procedure does not provide

that confirmation or no alternative evidence that is sufficient or appropriate is obtained. The audit team must consider significant limitations when planning further audit procedures and, when relevant, assessing the impact on the audit opinion.



The proposed logic shifts attention from how a document looks to whether its origin, transformation history, and relationship to the underlying economic event can be independently verified.

Figure 1. From Documentary Appearance to a Verifiable Chain of Trust

Source: Developed by the author.

Research Problem and Questions

The research problem is not only to find out if a document was written by an AI. The question is, under what circumstances such a document has sufficient evidential value for the purposes of an audit. The more the auditor can track the source and modification of a document, associate its assertions with observable financial or operational flows, and corroborate it by parties or records independent of the generation process, the more the document can support audit judgment. Reliance is also contingent on whether the output was reviewed and signed off by identifiable individuals and whether the entity had adequate governance and internal controls over its production [11], [12]. These conditions are mutually supportive. An external confirmation may confirm that a transaction occurred, but it may not say much about unauthorized modification. Strong internal governance cannot alone prove that the reported economic event occurred. Therefore, the problem requires an integrated assessment of the production history, accountability and connection to the underlying event of the document [8], [5], [9].

The main research question can be formulated like this: When used as audit evidence, the auditor should evaluate the reliability of the accounting and financial documents produced or materially altered by generative AI as follows: This inquiry is concerned with four subsidiary questions. First, what value does the ability to trace the provenance, the process of creation and later changes of a document, add to its evidential reliability? Second, how does corroboration through independent parties, cash flows, operational records or other evidence affect the auditor's ability to rely on the document? Third, what is the impact of oversight by humans, accountability, internal controls and AI governance on the reliability of documentary evidence? Fourth, if one or more of these conditions cannot be verified, what should the auditor do [8], [11], [9], [5], [13]?

Research Objectives and Significance

This study presents a sequential framework to convert the evaluation of the document generated or modified by AI to suitable audit response. The primary aim is to classify the document by its type and degree of AI involvement in its creation. The objective is to assess the evidential reliability using AGDAR six dimensions with specific focus on provenance and independent corroboration. Next, translate the evaluation into processes responsive to the particular source of uncertainty. If a document has poor provenance, weak traceability, or limited support from external sources of evidence, it may be necessary to adopt direct confirmation, physical inspection, or other extensive alternative procedures [8], [5]. On the other hand, if provenances are strong and there is a recorded human approval, protected edit history, and lasting independent evidence, then the need for further procedures will decrease, but it won't disappear altogether [9], [11]. According to the current AGDAR program, procedures are not automatically prescribed on the basis of a single reliability score. When a weakness is identified (for risks linked to the assertions based on the applicable audit framework), AGDAR will link it to an assertion-specific and risk-based audit response [2].

Considering the application of known principles of audit evidence takes place in a technological environment where documents can be created dynamically, the study seems crucial. ISA 500 gives some principles based directive for the auditor for evaluating the relevance and reliability of information including the source and the context in which these were obtained. The relevance of the principles remains strong, but their implementation will be increasingly more difficult in practice when ownership is shared between a user, a generative model, retrieved data, and human editing thereafter [8]. A well-known dichotomy is between internal and external evidence. Simply being an external document is not necessarily enough. For example, a document which seems to originate from outside may have been produced or reconstructed inside. Further, several apparently different documents may be derived from the same automated source. Reformatting and making language internally consistent tends to lower the diagnostic significance of appearance, but visual experience and professional judgment are still required [6], [7] AGDAR does not supersede ISA 500 nor does it create a separate evidential standard. A detailed analysis of provenance, integrity, traceability, independent corroboration, accountability, and consistency [9], [5] takes place according to contemporary reliability standards.

Methods

AGDAR was conceptually developed in four analytical stages. Phase one of the study comprised of a review of relevant academic literature, audit requirements, professional guidance and technical publications to identify risks of accounting and financial documentation which are mediated by AI. Greater attention was paid for issues recurring frequently, including the origin of information, unauthorized changes, reconstruction of the internal production process, external confirmation, human responsibility and consistency with underlying records [8], [2], [9]. Furthermore, risks were compared between sources and consolidated to reduce conceptual overlap. Thirdly, identical risks were categorized according to evidential function. The six dimensions of provenance, integrity, traceability, independent corroboration, accountability until consistency were led by this process. Consequently, the dimensions were extracted from the most frequent reliability issues rather than the original list [5], [7].

We also have a separate analytical step that classified the degree of AI participation in document production. The categorization depends on the level of control over the processes the human user has. The other variable is the level of independence with which the system adds to the end content. The materials reference documents made essentially by people with little AI intervention (eg formatting, linguistic structuring), documents primarily drafted or edited by AI but validated and documented by people, and documents principally generated independently, with no human verification or provenance. Just because people engage with new AI doesn't mean they're making more of those mistakes. When coupled with insufficient documentation, little human review or an unobserved generation trail, greater autonomy may instead magnify uncertainty [11], [3], [4]. Thus, each level is associated with a related audit response rather than an automatic accept/reject [8], [5].

Theoretical Framework and Prior Literature

Generative AI and the Changing Nature of Accounting Documents

One important difference is the use of AI in the audit process. Auditors use AI as an assistive tool, embedded into the audit workflow. AI may be used to structure information, detect patterns, locate relevant material or assist in performing procedures. The main concerns are the control of the tool by the auditor, the validation of the tool's output, confidentiality and the use of professional judgment [3], [4], [11] in that context. The evidential problem is different when a document is presented to the auditor which has been created or materially modified by the audited entity using AI. Here the technology is not just aiding the examination, it has helped create the information being examined. The auditor might not be aware of what data was fed into the model, how the output was edited, or if the source that appears on the document is its actual production history [5], [9]. AGDAR is mainly interested in the second scenario, that is, the framework that evaluates AI-shaped evidence, not the performance of an AI tool used by the auditor.

Generative AI's probabilistic operation creates a reproducibility issue for audit evidence. The solution to this task may differ in wording, explanation, value or document structure depending on the prompt, retrieved context, model version, system settings or conversation history. The same apparent inputs

might not produce the same outputs from the generation process. This means that if the last file is retained, the auditor may not be able to verify a document by reperformance in the normal way [2], [5]. This loss of reproducibility may interfere with the investigative trail, obscure whether a difference was due to bona fide regeneration or intentional manipulation, and complicate the determination of which interpretation or version was ultimately approved [14]. To analyze the problem, you need to preserve the original prompt, instructions, input data, retrieved sources, model and application versions, generation settings, timestamps, user identity, original output, any later revisions, and evidence of human review. Cryptographic hashes or electronic signatures can prove the state of a document at a certain time but cannot prove that its underlying content is economically correct [9], [7]. Taken as a whole, the records provide a generation trail that extends the traditional audit trail from the recorded transaction to the process that created the documentary content.

Results and Discussion

Audit Evidence and Information Reliability in a Digital Environment

Auditing is based on obtaining sufficient appropriate audit evidence to provide a basis for the conclusions on which the auditor's opinion is based and reliability is an important characteristic of the information used as audit evidence [8], [15], [5]. Recent professional developments reinforce the importance of assessing the relevance and reliability of information in a technology-driven environment rather than assuming a direct relationship between documentary form and strength of evidential value. The problem is no longer just a question of internal versus external evidence since the introduction of AI-generated documents. The auditor needs to consider who created the evidence, how the evidence is created through a technical process and how independent the generating system or party is from those with an interest in the outcome of the audit [11], [12].

The recent literature suggests that enterprise adoption of AI may alter traditional assumptions of evidence stability and integrity of audit trail [14], [13], [9]. Where content is generated or reformatted dynamically, the evidential context "behind the document" can include prompt history, sources retrieved, model versions, system logs, and human approvals. This supports a development of the traditional audit-trail idea into a generation trail that can re-construct the production of content rather than only tracing the posting of a transaction through an accounting system [5], [2].

Distinctive Risks of AI-Generated or AI-Modified Documents

Risks of AI-generated documents can be categorized into four interrelated groups. First, the risks of authenticity, where a document features a name, logo, signature, or format that indicates an origin that cannot be validated [12]. Second, integrity risks where content may have been altered, regenerated, or partially substituted without a trusted record of change [7]. Third, semantic risks are those that the document is coherent but contains a hallucinated fact, unsupported inference, or incorrect interpretation [2]. The fourth are accountability risks, where it is not clear who reviewed, approved or was responsible for the final information [5], [11]. These risks are increased when automated generation is coupled with poor documentation or an unstable change history [6].

The emerging problem of deepfaking financial documents demonstrates that, in addition to its semantic content, the visual architecture of evidence can be a target of fabrication [6], [7]. Logos, seals, signatures, templates, and document structures are so closely mimicked that visual cues are less useful as an initial basis for trust [2]. Therefore, audit verification aims at establishing the relationship between the document and the underlying economic event using independent sources such as external confirmation, system-generated traces, bank transactions, shipping records, device records or cryptographically verifiable provenance [9], [8].

Digital Content Provenance and Authenticity as Extensions of Audit-Evidence Reliability

Content provenance provides information for a user to understand the origin and history of a digital asset, and its interactions with the people, systems and tools that created or modified it [7], [6]. The C2PA Content Credentials architecture offers a technical framework to associate digital content with cryptographically verifiable assertions of the content's provenance and editing history [2]. NIST also discusses provenance mechanisms, including metadata, watermarking and other technical means, as possible ways to obtain information about whether content was generated or modified using AI, and to supplement recipient signals about origin and integrity [9].

From an audit perspective, Content Credentials should be seen as an additional level of trust rather

than a replacement of other audit evidence [7]. Even when a file is proven to be from a particular system it does not mean that the data behind it was economically valid or that the transaction happened as described [9]. However, the absence of standardized provenance information may not be a sign of fabrication per se, but rather adds to uncertainty and may require additional external corroboration, or alternative procedures [8]. This risk-based approach to accounting depends on professional judgment and skepticism as opposed to strict rules specific to technology [11], [10].

Professional Skepticism and Automation Bias

As polish, coherence, and visual persuasiveness increase, fluency and formatting may create cognitive confidence that is stronger than the evidence. Professional skepticism becomes even more important. Skepticism is recently conceptualized as a sensemaking process: noticing that something does not make sense and actively seeking information that is able to explain the inconsistency [10], [2]. So in a GenAI context the auditor needs to go beyond asking if a document 'looks right' to asking if there is an independent chain of evidence explaining why it should be true and what evidence would contradict it if it was generated or altered automatically [11], [12], [8].

AI content detectors do not replace professional skepticism [7]. Detection algorithms can produce false positives and false negatives, and generation techniques can move quickly and evade a single technical signal. A more robust audit response includes provenance information, metadata, system logs, independent sources, consistency with data in the ledger, linkage to the underlying economic event and analysis of the controls surrounding creation and approval [9], [8], [11], [5]. This is a more solid approach than depending on one single technical indicator that can be wrong or bypassed [6].

Research Gap

The existing literature contains many necessary bases to consider the problem at hand from another perspective. Research on large language models in accounting and auditing tends to ask questions such as how the artificial intelligence can help the professional by drafting, information retrieval, workflow support and audit task performance [1], [3], [4]. The emerging connection between AI, accounting, and finance is also being influenced by broader evaluations and research agendas [16], [17]. These contributions of audit firms show what auditors can do with AI. However, they are not as effective in the opposite context, in which the information provided to the auditor has been generated or materially altered by AI. The object of evaluation is not only the auditor's instrument of audit, but also the evidential chain behind the document.

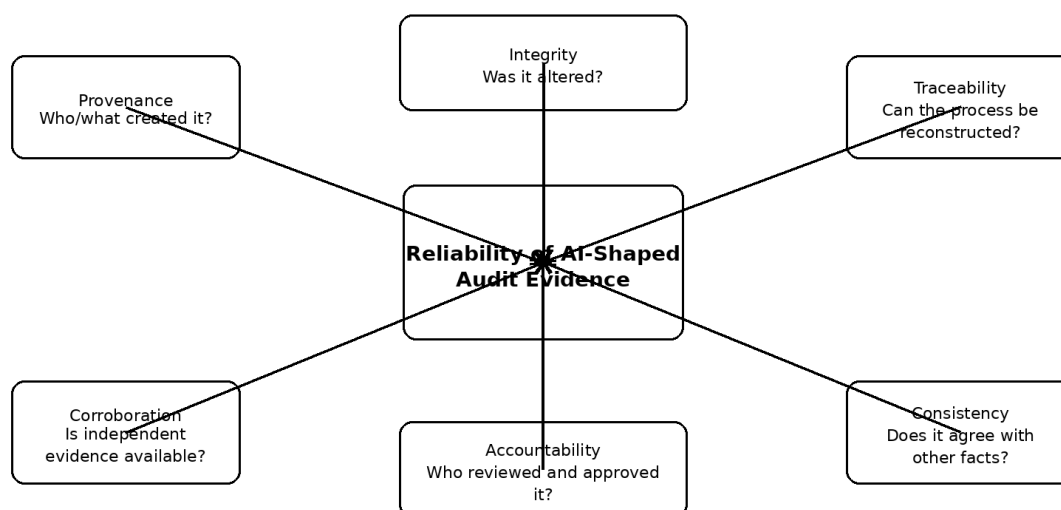
Related work involves the governance of artificial intelligence, synthetic content, traditional audit trail disruption and provenance. Recent research links generative systems to problems of evidence, accountability, and continual assurance [5], [14], [13]. Technical and legal contributions provide helpful insights on content credentials and deepfake financial documents [9], [7], [6]. Our contributions are not mere theoretical ones, but possess useful insights in different areas. Occasionally, provenance emerges in talks without being matched with any economic verification account. Otherwise some talks include discussion of governance without demonstration of its impact on reliance on a specific document. Lastly, authenticity is sometimes discussed as if equating evidential sufficiency. Nonetheless, none of these views by itself captures more than a piece of the credibility dilemma.

AGDAR takes care of this fragmentation by bringing together the right considerations for the life cycle of a document. Going by the process for origin creation and modification, and also responsibility and review preparation for verification with records and evidence outside of generation. Consequently, technology is less relevant than contribution. It encompasses a single evidential framework for provenance, integrity, traceability, corroboration, accountability and consistency. The problems in the structure are associated with the severity of the audit response [8], [11]. There is a difference between the binary test of whether a document is AI-generated and technical authentication models that do not look at the transaction underneath [9], [5].

The Proposed AGDAR Conceptual Framework

AGDAR was designed to capture the essential stages and connections in the life cycle of a document starting from initial creation, through modifications, reviews, approvals and final application as audit evidence. These six dimensions aren't just an arbitrary list or six independent pass/fail tests. Provenance gives information about where it came from, integrity asks was it altered without

authorisation, traceability reconstructs its history of production, independent corroboration tests its claims against other evidence outside of that process, accountability says who approved and used it, and consistency compares it to related records and known facts. Together, these dimensions connect the technical origin to the economic verification and professional responsibility [8], [9], [5]. The lifecycle perspective is further supported through research on synthetic-content risk, governance and deepfake documents [2], [13], [6].



The six dimensions are jointly assessed; no single dimension is a stand-alone pass/fail test.

Figure 2. The AGDAR Conceptual Framework

Source: Developed by the author based on the conceptual synthesis of the study.

Table 1. Dimensions of the Proposed AGDAR Framework

Dimension	Audit meaning	Examples of verification indicators
Provenance	Ability to establish the document's creator and the entity or system from which it originated.	Verified creator identity; digital signature; Content Credentials; independent source.
Integrity	Verification that no unauthorized modification occurred after creation or approval.	Hash; version history; timestamp; access and modification controls.
Traceability	Ability to reconstruct creation, modification, and review steps.	Prompt history; system logs; version history; audit/generation trail.
Independent corroboration	Evidence from other sources supporting the document's content.	External confirmation; bank data; shipping record; third-party evidence; reperformance.
Accountability	Clarity of human and organizational responsibility for the document and its approval.	Reviewer identity; authorization level; human sign-off; AI-use policies.
Consistency	Alignment of the document with records, facts, and other evidence.	Ledger matching; dates; quantities; contracts; analytical procedures.

Source: Developed by the author based on IAASB [8], C2PA [9], Autio et al. [2], Chandra et al. [7], Ahmad et al. [5] and Martinez [6].

Provenance

Provenance is the first anchor of AGDAR. Unless the document's origin is known, the auditor is unable to assess the evidence's source. In an AI-driven process, the source is more than just the name on the document. It mentions the individual or account that authored it, the model or application that was utilized, the data or retrieval resources that were furnished to it, as well as the individual who assessed the output that ensued [2], [5]. A cryptographic record, identity control and independent log can also strengthen the chain so that claims to origin are testable instead of merely visible [9], [7]. Origin is only the beginning, not the end A document might not always be based on accurate data or may describe an event not occurring [8].

Integrity

Integrity means whether a document has been changed since its creation or approval, and whether changes were permissible and traceable [9], [5]. In traditional environments, auditors might use access control, version management and system permissions; however, in GenAI environments, regeneration and partial rewriting can lead to substantive changes which are not visually obvious in the final version [11]. Thus, version histories, timestamps, digital hashes, signatures and controls that prevent or record modification are increasingly valuable, especially where documents directly affect recognition, measurement, presentation or disclosure [7].

Traceability

Traceability means being able to trace the document-production chain: inputs, prompts, sources of retrieval, version of the model or application, human reviews and subsequent changes [5], [11]. Since the output of GenAI may be attributed to a chain of interactions and not a single event of recording, the idea of a "generation trail" is presented in addition to the conventional audit trail [14]. This is especially true for retrieval-augmented generation, or AI agents that access multiple sources and take semi-autonomous steps. In these cases, keeping a record of workflow logic and technical path is important to understand the origin and reproducibility of information [4], [9].

Corroboration (Independency)

The second anchor is independent corroboration, testing the document from outside of its own production chain. Evidence is not truly independent if both the document and the material offered to support it originate from the same model, source data, or automated workflow. Instead, suitable corroboration may include direct confirmations, bank transactions, shipping or tax records, information obtained from third parties, or reperformance [8], [6]. This dimension tackles the lack of provenance, but at the same time, it puts into question documents with known origin: traceable generation cannot be proof that the underlying transaction is economically valid [9], [7]. If more AI involvement is desired or a less visible generation trail is desired, the auditor should rely more on evidence that is operationally separate from the document under review [5].

Liability

Accountability also specifies the person or organizational unit that created, reviewed and approved the document, thereby preventing the phrase "the system generated it" from creating a vacuum of responsibility. To govern AI in an effective manner, organizations need to identify process owners, data owners, human reviewers, and permissible levels of reliance [14], [11], [13], [12]. The risk of reliability increases when AI-generated documents enter accounting systems without a clear point of approval, because it becomes difficult to allocate responsibility for assumptions, errors and modifications [2].

Consistency

Consistency Testing whether the document is consistent with the larger body of records and facts. A good synthetic document may be a perfect match to a template, but not to transaction timing, quantities, contract terms, ledger entries, counterparty data, or larger business patterns [6]. Therefore, consistency testing amongst fields, systems and periods should be utilized by auditors as small inconsistencies are more likely to be revealing rather than obvious defects [8], [3], [5]. This finding is consistent with professional skepticism as an attempt to understand and explain evidence rather than merely checking on the form of the evidence [10].

Degree of AI classification involvement

The six dimensions are used in conjunction with the four levels of AI involvement, not as a scale of assumed truthfulness. A human document may be unreliable because either the source may be

biased, or controls may be weak. If the inputs, processing history, approval and external support are documented, a fully generated document may be usable. Instead, the classification measures the extent of uncertainty introduced by the method of production and the extent of visibility of the method to the auditor [2], [11]. Drawing a line between AI assistance and substantive generation also prevents minor drafting assistance from being conflated with unverified automated output [3]. The audit response should be based on the interaction of AI involvement, provenance, accountability and corroboration [9], [5] and not just on the technology label.

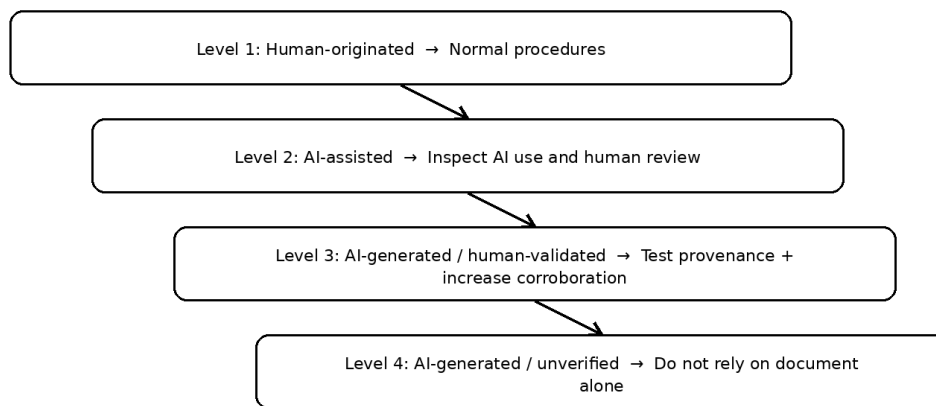
Table 2. Levels of AI Involvement and Proposed Audit Response

Level	Nature of document	Provenance risk	General audit response
1	Human-originated: prepared by a human without material GenAI involvement.	Conventional; depends on source and controls.	Normal audit procedures with assessment of source reliability.
2	AI-assisted: human is the principal author; AI assists drafting or organization.	Moderate where AI involvement is documented.	Inspect inputs and human review; obtain corroboration as needed.
3	AI-generated / Human-validated: AI creates substantive content followed by documented human approval.	Relatively high because of reliance on generation.	Test provenance, generation history, approval, and increase independent corroboration.
4	AI-generated / Unverified: automated generation without documented human verification or verifiable provenance.	Very high.	Do not rely on the document alone; obtain independent evidence and extensive alternative procedures.

Source: Source: Author, based on FRC [11], IAASB [8], Autio et al. [2], C2PA [9] and Ahmad et al. [5].

Logic for Applying the Proposed Framework

First the application identifies the document's role in "audit" procedure and the extent to which it was created with AI. In this instance, the inspector makes sure the six aspects comply with the relevant assertion rather than issuing the document one global rating [8], [5]. As a result, a trust gap emerges: the incongruity between the proof that auditors require and the verifiable features available. The gap may arise in the absence of origin records, changes that cannot be reconstructed, unclear responsibility, and inadequate external support. The last step does not involve automatic rejection but rather a proportionate response comprising additional corroboration, reperformance, direct confirmation or alternative procedure focused on the source of uncertainty [2], [13].



As AI involvement rises and traceability weakens, reliance on the document itself should decline while independent corroboration and alternative procedures should increase.

Figure 3. Graduated Audit Response to AI Involvement and Traceability

Source: Developed by the author.

This logic divorces quality of evidence from the means of production. A fully AI-generated document can be more powerful than a human-created document if the inputs and transformations used to generate it are captured, the responsibility for approval is clearly assigned, and the content of the document is in agreement with independent evidence [11], [9]. Or the reverse can be true: human authorship does not overcome a non-independent source, weak controls or inconsistency with the underlying transaction [8]. In this sense the six dimensions are complementary, although not necessarily weighted in the same way in each case. Provenance and independent corroboration are typically the primary anchors, but integrity, traceability, accountability and consistency determine if those anchors form a coherent and defensible chain of trust [7], [10].

Research Propositions

P1: The less the auditor can verify the origin, history of creation and modifications of an accounting or financial document, the less reliable it is as audit evidence [6], [8], [5]. The proposition is grounded in the assumption that the lack of provenance raises the level of uncertainty about whether a document is a representation of the event it pretends to represent. At the same time, verifiable traceability provides meaningful trust signals without replacing verification of underlying economic event [9], [7].

P2: A document generated by AI is not inherently unreliable; reliability is determined by provenance, integrity, traceability, corroboration, accountability, and consistency [15]. This proposition is needed to avoid binary judgment and mirrors professional approaches that focus on output quality, governance, controls and human judgment instead of banning a technology just because it is generative [11], [2], [3], [13].

P3: The greater the use of GenAI in content creation, or the more difficult it is to trace the generation chain, the greater the need for independent corroborative evidence [5], [10], [11]. Weaker direct signals from the document itself will need to be compensated by external confirmation, reperformance or other independent evidence, as automated generation may make it more difficult to understand the source of information in the final document [8], [6].

P4: Mechanisms for verifiable content-provenance enhance the auditor's ability to establish the authenticity and integrity of documents, but do not replace the verification of the underlying transaction [7], [8], [6]. The proposition weighs the technical value of cryptographic provenance and origin records with their epistemic limitations: a technically valid chain may stem from inaccurate or fraudulent source data and provenance must, therefore, be supplemented with independent substantive evidence [9].

P5: The uncertainty of the person who prepares, reviews or approves the document may increase the risk of the audit relying on a document [11], [14]. As tools have developed generative or agentic capabilities, a fundamental tenet of AI governance is clear roles and responsibilities, with human accountability [12], [2], [13].

P6: Professional skepticism is more effective in a synthetic documents environment when it moves away from looking at the form and instead focuses on testing explanations, consistency and contradictory evidence [10], [8], [7] Active searches for independent sources and for information that contradicts the apparent narrative of a document are likely to be more useful than attempts merely to determine whether the writing style 'looks like AI' [6], [5].

Discussion and Professional Implications

Ultimately AGDAR sets out an orderly interrogation into the evidential chain of documents rather than an impressionistic reading of documents. An invoice, contract or memorandum that is presented in a neat format may give off the impression of credibility. It does not tell you how it was created, however. The framework directs the auditor to identify specific interruptions to this flow: unidentified origin, unrecorded alteration, unauthorized approval, proof originating from the same system, and lack of agreement with transaction records [8], [5].

Each of these deficiencies will require a different kind of response. Whenever a provenance is strong, we can reduce uncertainty but when it is weak we get that it is better to redirect attention to direct verification and other independent evidence [9], [11]. The contribution does not guarantee the detection of all error or fraud, but it is a method for reducing the risk that either will be hidden behind documentation [2].

AGDAR also identifies a control question that is likely missed by routine document review. The mere authorization of an accounting entry does not imply the control over the contents which would influence an accounting entry. Entities should therefore consider who can initiate AI generation, what models and data can be used, how interactions are retained and who is responsible for the final document [11], [12]. As highlighted by Autio et al. [2] and Nandal [13], system compliance entails that instead of being optional technicalities, identity management, generation logs, source validation, version controls and documented approval become relevant parts of the control environment. According to the novel research results, the present position remains human accountable. It is consistent with the adaptive and agentised auditing environment [14], [4].

The framework takes the identified fragility and translates it into a response at the level of procedure. If no origin information is available, verification may be done directly. Uncertainty about changes may lead to checking logs or the initial file. Documentation proving review and approval may be required when there is no control evidence. Reperformance or cross-system comparison might be needed for inconsistency. Metadata and content credentials can help with origin and integrity, but neither should be relied upon for transaction proof [9]. According to FRC [11] and Martinez [6], similar to an audit, an AI-content detector is not a replacement of audit evidence or professional judgment but a technical signal. The procedures should be suitable for the assertion and matching with the identified trust gap [8], [10].

Conclusions and Recommendations

The research highlights the issue of generative AI that doesn't just involve synthetic authorship. It creates an audit problem that revolves around an evidential chain of which may become incompletely or completely unknown. Such a chain violates the need to assess the source and reliability of an information. AGDAR does a reply by tracking a document from formative, through transformation, review and approval, through comparison with outside evidence. The visual credibility does not suffice and no one technical signal can determine reliability, according to the six dimensions [8], [9]. The main cornerstone of provenance and independent corroboration is integrity. Traceability, accountability, and consistency are three other key anchors that make the document's environmental life cycle testable [2], [5], [6], [11].

The framework also doesn't group all AI-generated documents together. An independently supported output that is traceable may be more convincing evidence than a human produced document where there are weak controls. On the other hand, if data is wrong or fake, technical proof of origin cannot generate economic truth [9], [7]. Thus, the answer is multilayered. Technical verification refers to everything we know about production, professional skepticism refers to the explanation surrounding the document, and independent evidence refers to linking its claims to the underlying event [8], [10]. This structure offers a stronger foundation for judgment than either automatically accepting or automatically rejecting results.

Recommendations

Audit firms should implement clear policies regarding any document that may be created or substantially revised by GenAI. The audit methodology should pose direct inquiries about the source, history of development, model or application involvement, as well as documented human review. These assessments will relate to the analysis of the risk of material misstatement [8] [12] [2].

It is essential for training to encompass material provenance, metadata, limitations of synthetic-content detectors and the enduring principle that professional judgment and responsibility remain that of the human auditor [11], [9].

Standard setters and professional bodies might develop additional application guidance for evaluating information from generative or hybrid systems, including for circumstances where source information is untraceable or inconsistent with other evidence [8], [9]. We should push for a bigger interoperability among audit standards and digital-provenance specifications so they can not insulate technical authentication mechanisms from professional evidential reasoning while at the same time retaining the principle-based standards to allow rapid technological changes [15], [7], [11].

Future research aimed at AGDAR can rely on empirical investigations to test our theory using a variety of methods. These can include behavioral experiments which manipulate the degree of AI involvement and the availability of provenance information, case studies of organizations adopting GenAI tool, or studies of whether dense digital signals produce new forms of automation bias [16], [10], [5]. Thus, the current framework establishes a theoretical grounding that may subsequently be transformed into hypotheses, constructs and measurements without the current study having a field component [17], [4], [13].

Limitations

The framework is theoretical and has not yet been field tested, experimentally tested or tested with archival audit evidence. The practical application is also somewhat related to the information-technology infrastructure of the audited entity. The system logs, version histories, identity and access records, documentation on AI use, and evidence of human approval(i.e. audits of algorithms and impact assessments) could provide more visibility into providence, integrity, traceability and accountability. AGDAR is still relevant in the absence of any such records however, the gap of trust will be bigger and the auditor will require further independent corroboration or alternative procedures [8], [11]. As a result, the output of an application may differ depending on the digital maturity and control environment of an entity [2], [9]. Another limitation is the evolving nature of generative technologies and professional advice, including the exposure-draft status of the 2026 ISA revisions. In actual practice, there is the need to engage in an empirical study of the weights that auditors will assign to all six dimensions [17].

References

- [1] H. Li and M. A. Vasarhelyi, "Applying large language models in accounting: A comparative analysis of different methodologies and off-the-shelf examples," *Journal of Emerging Technologies in Accounting*, vol. 21, no. 2, pp. 133–152, 2024, doi: 10.2308/JETA-2023-065.
- [2] C. Autio, R. Schwartz, J. Dunietz, S. Jain, M. Stanley, E. Tabassi, P. Hall, and K. Roberts, *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile (NIST AI 600-1)*. Gaithersburg, MD, USA: National Institute of Standards and Technology, 2024, doi: 10.6028/NIST.AI.600-1.
- [3] H. Gu, M. Schreyer, K. Moffitt, and M. Vasarhelyi, "Artificial intelligence co-piloted auditing," *International Journal of Accounting Information Systems*, vol. 54, Art. no. 100698, 2024, doi: 10.1016/j.accinf.2024.100698.
- [4] F. Xiong, Q. Han, and C. Zhang, "Design AI agent for auditing: Applying large language models (LLMs) and retrieval augmented generations (RAG) to audit workflows," *Journal of Emerging Technologies in Accounting*, vol. 23, no. 1, pp. 189–198, 2026, doi: 10.2308/JETA-2024-041.
- [5] S. Ahmad, A. Harbola, S. Sachdeva, B. K. Yadav, M. K. Sharma, A. Khan, and J. Ahmad, "Generative AI in enterprise IT implications for audit evidence, assurance, and governance," *EDPACS*, advance online publication, 2026, doi: 10.1080/07366981.2026.2633846.

- [6] A. L. Martinez, "Jurisvision of deepfake financial documents: Semiotic disruption and remediation in post-textual evidence," *International Journal for the Semiotics of Law*, advance online publication, 2026, doi: 10.1007/s11196-026-10514-w.
- [7] B. Chandra, J. Dunietz, K. Roberts, Y. Lee, P. Fontana, and G. Awad, *Reducing Risks Posed by Synthetic Content: An Overview of Technical Approaches to Digital Content Transparency (NIST AI 100-4)*. Gaithersburg, MD, USA: National Institute of Standards and Technology, 2024, doi: 10.6028/NIST.AI.100-4.
- [8] International Auditing and Assurance Standards Board, "Proposed revisions for audit evidence & risk response: ISA 330, ISA 500 & ISA 520," Exposure Draft, 2026. [Online]. Available: <https://www.iaasb.org/publications/proposed-revisions-audit-evidence-risk-response-isa-330-isa-500-isa-520>
- [9] Coalition for Content Provenance and Authenticity, "C2PA specifications, version 2.4: Content Credentials," 2026. [Online]. Available: <https://spec.c2pa.org/specifications/specifications/2.4/specs/ContentCredentials.html>
- [10] G. Xu and C. Yang, "It doesn't make sense: Professional skepticism as sensemaking," *Current Issues in Auditing*, vol. 19, no. 1, pp. P10–P15, 2025, doi: 10.2308/CIIA-2024-001.
- [11] Financial Reporting Council, "Generative and agentic AI guidance," 2026. [Online]. Available: <https://www.frc.org.uk/library/standards-codes-policy/audit-assurance-and-ethics/guidance/ai-in-audit/>
- [12] Public Company Accounting Oversight Board, "Staff update on outreach activities related to the integration of generative artificial intelligence in audits and financial reporting," 2024. [Online]. Available: <https://pcaobus.org/news-events/news-releases/news-release-detail/pcaob-staff-shares-observations-from-outreach-on-use-of-generative-artificial-intelligence-in-audits-and-financial-reporting>
- [13] N. Nandal, "AI-governed audit assurance: A conceptual framework integrating predictive intelligence, governance controls, and continuous assurance," *EDPACS*, advance online publication, 2026, doi: 10.1080/07366981.2026.2667950.
- [14] S. Ali, K. Mohiuddin, S. Sachdeva, A. Juyal, S. Praveen, and A. Khan, "Adaptive audit frameworks for autonomous and AI-assisted decision ecosystems," *EDPACS*, advance online publication, 2026, doi: 10.1080/07366981.2026.2660882.
- [15] International Auditing and Assurance Standards Board, "Technology position," 2024. [Online]. Available: <https://www.iaasb.org/publications/technology-position-statement>
- [16] M. M. Dong, T. C. Stratopoulos, and V. X. Wang, "A scoping review of ChatGPT research in accounting and finance," *International Journal of Accounting Information Systems*, vol. 55, Art. no. 100715, 2024, doi: 10.1016/j.accinf.2024.100715.
- [17] T. C. Stratopoulos and V. X. Wang, "Artificial intelligence and accounting research: A framework and agenda," *International Journal of Accounting Information Systems*, vol. 56, Art. no. 100760, 2025, doi: 10.1016/j.accinf.2025.100760.