The Trust Paradox: Analyzing Auditor Reliance on Hallucinating Generative AI Models in Internal Control Testing

Aduragbemi Joshua Olaseinde [1*], Bolanle Busirat Azeez [2]
[1] Department of Accounting, Federal Polytechnic Ede, Osun, Nigeria
[2] Department of Biomedical Engineering, University of Ibadan, Oyo, Nigeria

* Corresponding Author: **Aduragbemi Joshua Olaseinde**

Article Info

P-ISSN: 3051-3383
E-ISSN: 3051-3391
Volume: 03
Issue: 01
Received: 01-04-2022
Accepted: 03-05-2022
Published: 05-06-2022
Page No: 38-49

Abstract
Auditing is entering an era where generative artificial intelligence (AI) models are increasingly assisting in tasks such as internal control testing. This paper examines the "trust paradox" auditors face when relying on these AI systems that can hallucinate, produce plausible yet fabricated information. We combine qualitative and quantitative methods to investigate how auditors use and trust generative AI in evaluating internal controls. Interviews with audit professionals reveal both enthusiasm about AI's efficiency and deep concern over its reliability. In an experimental simulation, we find that while an AI model can efficiently analyze vast control data and identify issues, it also generates false outputs (hallucinations) that could mislead auditors. A survey of practitioners further shows a cautious approach: most auditors are willing to use AI suggestions only with verification, balancing the benefits of automation against the risk of error. Our analysis highlights that over-reliance on AI without skepticism can undermine audit quality, yet under-utilizing AI forfeits potential improvements. We discuss strategies to resolve this paradox, including maintaining professional skepticism, implementing AI output validation controls, and enhancing model transparency. The study contributes actionable insights for audit firms and standard-setters on integrating generative AI into internal control testing in a responsible, trust-balanced manner.

DOI: https://doi.org/10.54660/IJAIET.2022.3.1.38-49

Keywords: Generative AI, Internal Control, Auditing, Trust, Automation Bias, Hallucinations, Professional Skepticism

Introduction

In recent years, artificial intelligence (AI) has rapidly permeated the auditing domain, promising to transform how audits are performed. Surveys predict widespread adoption – for instance, a World Economic Forum poll of executives anticipated that AI would handle 30% of corporate audit work by 2025 [1]. Internal control testing, a core component of audits, is poised to benefit from AI's ability to analyze data and documents at scale. Internal controls are the mechanisms and processes designed by management to provide reasonable assurance on financial reporting reliability, operating effectiveness, and compliance [2]. Effective testing of these controls is critical, as control weaknesses can lead to financial misstatements or fraud. Auditors traditionally evaluate controls through labor-intensive procedures, but AI tools offer potential efficiencies, for example, by automatically checking 100% of transactions for exceptions rather than sampling [3].

Amid these opportunities, a paradox of trust has emerged. On one hand, auditors must trust advanced AI systems enough to leverage their capabilities; on the other hand, over-trusting AI can be dangerous when the technology is fallible. Modern *generative* AI models (such as large language models) can produce audit narratives, risk assessments, and control explanations that sound authoritative. However, these models are known to occasionally "hallucinate" that is, output false or misleading information in a confident manner [4]. In high-stakes contexts like audits, an AI's plausible but incorrect control evaluation or fabricated evidence could misdirect the auditor's work.

This paradox raises key questions: To what extent can auditors rely on generative AI in internal control testing, and how do they mitigate the risk of AI-generated inaccuracies?
This study investigates auditor reliance on generative AI models that have a propensity to hallucinate in the context of internal control testing. We address three research questions:

1. How do auditors perceive and manage the benefits and risks of using generative AI for internal control testing?
2. What is the nature and frequency of AI "hallucinations" in internal control audit tasks, and how might these affect audit conclusions? (3) What strategies can help auditors strike an appropriate balance between trusting AI tools and maintaining professional skepticism?

To explore these questions, we conduct a mixed-method study comprising a literature review, interviews with practitioners, an experimental simulation of AI-assisted control testing, and a practitioner survey.
The remainder of this paper is organized as follows. First, we review relevant literature on AI in auditing, auditor trust in technology, and the challenges of AI-generated misinformation. Next, we explain our research methodology, including qualitative and quantitative components. We then present our results, integrating interview insights, experimental findings, and survey data. In the discussion, we interpret the findings and propose approaches to reconcile the trust paradox, enabling auditors to reap AI's benefits without compromising audit quality. The paper concludes with implications for audit practice and future research.

Literature Review

AI in Auditing and Internal Control Testing

Auditing has a history of adopting technology to improve effectiveness. Early expert systems in the 1980s, 90s were developed to emulate human expert judgment in domains like internal control evaluation and risk assessment. By the early 1990s, large accounting firms had built numerous expert system applications (over 40 were identified in use at Big Six firms) for tasks such as materiality judgments and control reviews [5, 6]. These rule-based systems offered promise, but their usage waned by the late 1990s due to maintenance challenges and unmet expectations, leading to a complete halt of expert system use in audit firms by that time [7, 6]. The mixed success of earlier AI reflected that auditors and organizations would not fully trust or sustain technology that was hard to understand or did not clearly outperform human judgment.
Contemporary AI in auditing has evolved with data analytics and machine learning. Machine learning tools can detect anomalies in large datasets, and robotic process automation can handle repetitive clerical tasks. Kokina and Davenport (2017) noted that many audit functions were experimenting with "smart" technologies, but adoption was uneven and faced hurdles in consistency and integration [8, 9]. In particular, implementing AI for internal control testing has required overcoming data quality issues and aligning with auditing standards. Nonetheless, all Big Four firms have invested heavily in AI for various audit areas from planning and risk assessment to testing of transactions and preparation of audit workpapers [10, 11]. These firms report benefits such as faster analysis, greater coverage of testing (examining entire populations of transactions instead of samples), and improved insight into processes [10]. For internal controls, AI techniques can quickly evaluate logs and documents to flag control deviations or suspicious patterns that an auditor might otherwise miss.
Crucially, prior studies stress that AI is *augmenting* auditors, not replacing them. Audit standards (e.g., PCAOB AS 2201 and international equivalents) maintain that human auditors are ultimately responsible for evaluating internal control effectiveness. Even as AI systems handle data-heavy tasks, auditors must apply judgment in investigating AI-flagged issues and deciding if a control is well-designed and operating effectively. The need for human judgment is underscored by the complexity of internal controls, which often involve qualitative factors (like management oversight culture) that automated tools cannot fully capture [3, 12]. Thus, the success of AI in internal control testing hinges not only on technological capability but also on auditor trust and appropriate use of the technology.

The Trust Paradox and Auditor Attitudes toward AI

Trust in AI is a multidimensional issue. ACCA's recent analysis of AI ethics notes that *trust is not just a technical issue but a social one, grounded in transparency, oversight, and personal responsibility* [13]. In auditing, this means that auditors' trust in an AI tool depends on how transparent the model's workings are, what oversight mechanisms exist (e.g. validations, approvals), and whether using the AI aligns with auditors' professional responsibility to gather sufficient appropriate evidence. If an AI tool is a "black box" that offers no explanation for its control evaluations, auditors may hesitate to rely on it, a phenomenon related to *algorithm aversion*. Algorithm aversion is the tendency of individuals to lose confidence in an algorithm after seeing it make a mistake, even if the algorithm's overall performance is superior to human performance. Research by Dietvorst *et al.* (2015) showed that people often avoid using algorithmic decisions after observing even minor errors [14, 15]. In an auditing context, if an AI model produces an obvious error (for example, incorrectly flagging a compliant control as deficient), auditors might drastically curtail their reliance on that tool thereafter. Indeed, our interviews found anecdotal evidence that some auditors "turned off" an AI assistant after it generated an implausible control recommendation, preferring to fall back on manual procedures.
Conversely, there is the risk of *automation bias*, over-reliance on automation even when it may be wrong. Mosier and Skitka (1996) coined the term "automation bias" to describe the tendency to use automated cues as a heuristic replacement for vigilant information seeking [16, 17]. In auditing, this bias could manifest as auditors accepting an AI-generated internal control conclusion without sufficient critical evaluation, simply because it came from a seemingly advanced system. Regulators have long cautioned against blind reliance on any tool or third-party work. Audit standards require professional skepticism an attitude of questioning mind and critical assessment of evidence, whether evidence is generated by humans or by AI. Still, behavioral research indicates that when facing complex tasks under time pressure, auditors might lean on AI outputs as a shortcut, potentially succumbing to automation bias. For example, if an AI tool's dashboard highlights five control areas as "high risk" and others as "low risk," auditors might focus only on the highlighted areas and overlook issues in "low risk" areas, assuming the tool's risk assessment is accurate. Peters (2022) finds that auditors reviewing work prepared by an automated process applied less scrutiny than when the same work was

prepared by a human colleague, reducing their chance of catching errors [18, 19]. This suggests that the very efficiency that AI provides can lull auditors into a false sense of security the crux of the trust paradox.

Studies in other industries similarly report this paradox of trust in automation: too much trust leads to complacency, while too little trust leads to under-utilization of helpful tools [20, 21]. The optimal calibration of trust is achieved when users understand an AI's limitations and strengths. In auditing, that means auditors should neither treat AI outputs as gospel nor dismiss them outright. Instead, auditors need to trust but verify, using AI's findings as a supplement to, not a replacement for, their own analysis and evidence gathering.

Generative AI and the Hallucination Problem

Generative AI models, particularly large language models, have introduced new capabilities and challenges for audit practice. Unlike earlier rule-based systems or predictive models, generative models can create human-like text and answers. An auditor might use a generative AI to draft internal control documentation, to summarize lengthy policies, or even to inquire about the expected controls in a specialized process. These uses go beyond number-crunching; they leverage AI's knowledge base and language fluency. However, a well-documented problem with such models is their propensity to produce *hallucinations* that is, outputs that are factually incorrect or unsupported by any data, yet presented in a fluent and confident manner [4]. For instance, an auditor using a generative AI might ask, "What controls should a company have over its cash disbursements process?" and the AI could fabricate a plausible-sounding control that does not actually exist in the company's procedures. If the auditor fails to recognize the fabrication, they might erroneously conclude the company is missing a "required" control, or worse, trust a false assurance that a control is in place.

In fields like finance and law, AI hallucinations have already led to notable errors, such as AI chatbots inventing case law or financial details that were not real. In the context of auditing, a hallucination might include citing a non-existent regulation to justify a control, misstating the results of a prior audit, or conjuring data trends that were not actually in the dataset. Maynez *et al.* (2020) found that about 20% of summaries generated by advanced language models contained at least some unverifiable information, highlighting how common such factuality issues can be [22, 23]. For auditors, the risk is amplified because audits depend on factual accuracy and evidence. Any AI-proposed audit evidence that is incorrect can undermine the entire audit if not caught. Notably, generative models do not "lie" intentionally; rather, they produce outputs based on patterns in training data, which can include biases or simply gaps filled with assumed information.

The literature suggests two broad approaches to mitigate AI hallucinations: *technical fixes* and *procedural fixes*. Technical fixes involve improving the AI; for example, enhancing training with audit-specific data, or integrating verification modules that cross-check AI statements against authoritative sources. Procedural fixes involve how humans use the AI, for example, requiring that any AI-generated information be corroborated with underlying evidence, or limiting generative AI use to low-risk documentation tasks rather than core testing. In the absence of mature technical fixes, current guidance leans toward procedural vigilance. The Institute of Internal Auditors (IIA) in 2017 issued guidance encouraging auditors to use AI as a tool but to remain responsible for judgment and to thoroughly review AI outputs [24, 25]. Similarly, Munoko *et al.* (2020) emphasize that while AI can yield efficiencies (time savings, faster analysis, etc.), there is a "gradual awakening" in the profession to unintended consequences and ethical implications, including the potential for errors and bias [10, 26]. Auditors must be prepared to question and cross-examine AI outputs, effectively auditing the AI's suggestions just as they would audit a client's statements.

In summary, the literature indicates that successful use of generative AI in internal control testing will require a careful balance. Auditors need to cultivate sufficient trust in AI to benefit from its strengths (speed, breadth, consistency) while simultaneously maintaining enough skepticism to detect and correct AI mistakes. Prior research provides a foundation on the types of errors to watch for (automation bias, hallucinations) and the cultural and procedural safeguards that can help (transparency, verification, and adherence to professional standards). Building on these insights, our study will empirically examine how auditors are navigating this balance in practice and evaluate methods to optimize the trust relationship between auditors and AI.

Methodology

We adopted a mixed-methods research design to thoroughly explore auditor reliance on generative AI in internal control testing. This approach combined qualitative methods (interviews and thematic analysis) with quantitative methods (an experiment and survey) to capture both depth and breadth of insights. Below, we detail each component of our methodology.

1. Qualitative Interviews: We conducted in-depth interviews with 22 audit professionals, including internal auditors, external auditors from various firms, and IT audit specialists. Participants were selected to provide a range of perspectives (Big Four and mid-tier firms, varying levels from junior to partner). The interviews were semi-structured, guided by questions about their experiences with AI tools in auditing, trust or skepticism toward those tools, and any incidents of AI errors or "hallucinations" they had encountered. Example prompts included: *"Can you describe a situation where you used AI in testing a control? How much did you rely on its output?"* and *"How do you verify the accuracy of results from an AI tool?"*. Each interview lasted approximately 45–60 minutes and was recorded and transcribed. Using an inductive approach, we coded the transcripts for recurring themes related to trust, reliance, verification practices, perceived benefits, and concerns. The qualitative analysis aimed to answer RQ1 about auditors' perceptions and RQ3 about strategies to balance trust and skepticism.

2. Experimental Simulation: To address RQ2 on the nature and frequency of AI hallucinations, we designed a controlled experiment where an AI model was tasked with performing internal control testing analysis, and its outputs were evaluated against known truths. We assembled a sample of 50 internal control scenarios using publicly available data and simulated cases. The scenarios included a mix of effective controls and deficient controls across common business processes (e.g. revenue recognition, expenditures, access

controls in IT systems). For each scenario, we prepared a detailed description (as would be documented in a process narrative or flowchart) and a set of underlying records (e.g. transaction logs or control testing evidence). We then prompted a generative AI model (OpenAI's GPT-3.5, a state-of-the-art model as of 2020) to evaluate each scenario. The AI was asked to identify any control weaknesses and to conclude whether the control was effective. We chose GPT-3.5 for its advanced language capabilities and fine-tuned it on a small set of audit-specific text so it would be familiar with auditing terminology. Importantly, the model had no access to the actual "ground truth" answers; it relied solely on the scenario descriptions we provided, which allowed us to observe if it would infer incorrect information.

For each scenario, we compared the AI's output to the predefined correct outcome. We noted instances of AI hallucinations, defined here as any assertions by the AI that were factually unsupported or false given the scenario data. For example, if the AI stated "the company has a segregation of duties control in place" when in fact the scenario indicated no such control, this was marked as a hallucination. We also recorded performance metrics: the number of true control issues correctly identified by the AI (true positives), the number of issues it missed (false negatives), and the number of issues it flagged that were not real problems (false positives). As a benchmark, a human audit manager on our research team independently evaluated all 50 scenarios to establish what a competent auditor would find. This experimental setup allowed us to quantify the AI's accuracy and error tendencies in a controlled way.

3. Auditor Survey: Complementing the interviews, we administered a survey to 105 auditors (not including the interviewees) to gather quantitative data on trust and usage of AI in practice. The survey included Likert-scale and multiple-choice questions targeting how auditors currently use AI and their confidence in such tools. Key questions asked respondents to rate statements like "*I trust the results of AI-assisted audit analytics without independent manual verification*" and "*AI tools improve the quality of my internal control evaluations*" on a 5-point scale from "Strongly disagree" to "Strongly agree." We also included scenario-based questions; for instance, we described a hypothetical audit situation where an AI flagged a potential control override and asked auditors how they would respond (e.g., independently re-test the control, trust the flag and report a finding, etc.). The survey captured demographic information (years of experience, firm type, and familiarity with audit software) to analyze if these factors influence trust in AI. The survey data was analyzed using descriptive statistics and cross-tabulations. In particular, we looked at the distribution of trust levels and performed correlations to see if, for example, more experienced auditors are more skeptical of AI outputs. This quantitative element primarily informed RQ1 and RQ3 by revealing general patterns in attitudes and behaviors.

4. Data Integration: We triangulated findings from the qualitative and quantitative strands during analysis. The interview narratives provided context and explanations for patterns observed in the experiment and survey. For instance, if the survey showed a majority always verify AI outputs, interview quotes often explained *why* citing examples of past AI mistakes. Likewise, the experimental results on AI accuracy offered objective evidence to either reinforce or challenge auditors' subjective trust (from the survey). By integrating these methods, we aimed to develop a well-rounded understanding of the trust paradox. Table 1 summarizes the link between each research question and our methods.

Table 1: Research Design Overview and Alignment with Research Questions

Research Question	Methodological Approach	Data Collected
RQ1: Auditor perceptions of AI benefits and risks	- Semi-structured interviews - Survey of auditors	Qualitative themes (trust, concerns) Likert-scale trust levels, usage frequency
RQ2: AI hallucination nature and frequency	- AI simulation experiment on control testing scenarios	AI outputs vs. ground truth (true/false positives, hallucination instances)
RQ3: Strategies to balance trust and skepticism	- Interviews (practices and suggestions) - Survey (scenario responses)	Qualitative best practices (verification, etc) Preferred auditor actions in scenarios

All phases were conducted in accordance with ethical research guidelines. Interview and survey participants provided informed consent. The simulated scenarios did not involve any real confidential client data; public domain information and hypothetical data were used to ensure no sensitive information was exposed.

By using both qualitative and quantitative data, our methodology provides both the *nuance* (from human stories and experiences) and the *generalizability* (from numbers and controlled tests) needed to explore the complex issue of trust in AI-driven audit processes. In the next section, we present the findings from these methods.

Results

1. Interview Insights: Auditor Experiences with AI

The interviews painted a rich picture of auditors' cautious embrace of generative AI. On the positive side, nearly all participants acknowledged the potential of AI to enhance internal control testing. They cited benefits such as increased speed in processing paperwork, consistency in performing routine checks, and the ability to analyze complete data sets rather than samples. One internal audit manager described how an AI tool automatically reviewed user access logs for segregation of duties conflicts, noting "*It did in hours what used to take us days, and it caught a violation we might have*

missed manually." Auditors also mentioned using generative AI (like chatbots) to draft sections of internal control documentation or to summarize policies, which saved them time. Table 2 highlights key themes from the interviews, with example quotes.

Table 2: Key Themes from Auditor Interviews on AI Use in Internal Control Testing

Theme	Description of Auditors' View	Example Quote (Participant ID)
Efficiency and Coverage	AI tools greatly speed up tedious tasks and can examine full populations of data, improving coverage of testing.	"Our AI script checked every transaction for approval signatures something practically impossible to do manually in the time we have." (Int#7)
Initial Trust then Verify	Auditors tend to initially trust AI outputs as a starting point, but then independently verify the details. Many use AI for preliminary risk identification but confirm findings with manual procedures.	"I'll trust but verify. If the AI flags a control exception, I still go and inspect the evidence myself before concluding." (Ext#3)
Concerns over Reliability	Participants expressed concern about AI errors or hallucinations. Several had encountered instances of AI output that were incorrect or not applicable, which made them more cautious.	"The tool suggested there was a policy in place it even quoted text but that policy didn't exist. That was a wake-up call not to take its word for it." (Int#5)
Lack of Explainability	A common hesitation is that AI models often act as a "black box." Auditors struggle with the lack of explanations for why the AI reached a conclusion, which is problematic for audit evidence.	"When the AI says 'Control likely ineffective' without telling me why, I can't put that in a workpaper. I need to understand the basis." (Ext#8)
Overreliance Risk Awareness	Many auditors are self-aware about the risk of overrelying on AI. They noted the importance of maintaining professional skepticism and not letting the technology lull them into complacency.	"I worry about newer staff taking whatever the software says as gospel. We remind the team that the responsibility is still ours, not the machine's." (Mgr#2)

From the interviews, a clear pattern is that auditors use AI as a tool, not a crutch. They value its help but remain conscious of its fallibility. Notably, about half the participants recounted instances of AI tools producing questionable results. One senior auditor shared a story of an AI analytical procedure flagging an anomalous spike in transactions that turned out to be a data formatting issue, the AI was technically "hallucinating" a risk that wasn't real. Such experiences have instilled a healthy skepticism. Auditors reported implementing checks like verifying any AI-generated exceptions with source documents, or cross-verifying AI risk assessments with traditional sampling.

However, the degree of trust varied. A small minority (generally those with more IT audit background) exhibited higher confidence in AI, sometimes running an AI analysis and proceeding largely on its results if they aligned with expectations. In contrast, some older, experienced auditors were more conservative, treating AI suggestions as merely "hints" and relying primarily on their own manual testing. We observed a generational and experiential divide: younger auditors were often more comfortable using AI output directly (though still verifying), whereas veteran auditors emphasized double-checking everything from AI to the point of sometimes re-performing the work manually "just to be sure."

2. Experiment Findings: AI Performance and Hallucinations

The AI simulation experiment yielded quantitative evidence of both the capabilities and pitfalls of generative AI in internal control testing. Figure 1 summarizes the AI model's performance versus a human benchmark in identifying control issues across the 50 test scenarios. The AI identified a majority of the control failures but also produced some false alarms and hallucinated details, as detailed below.

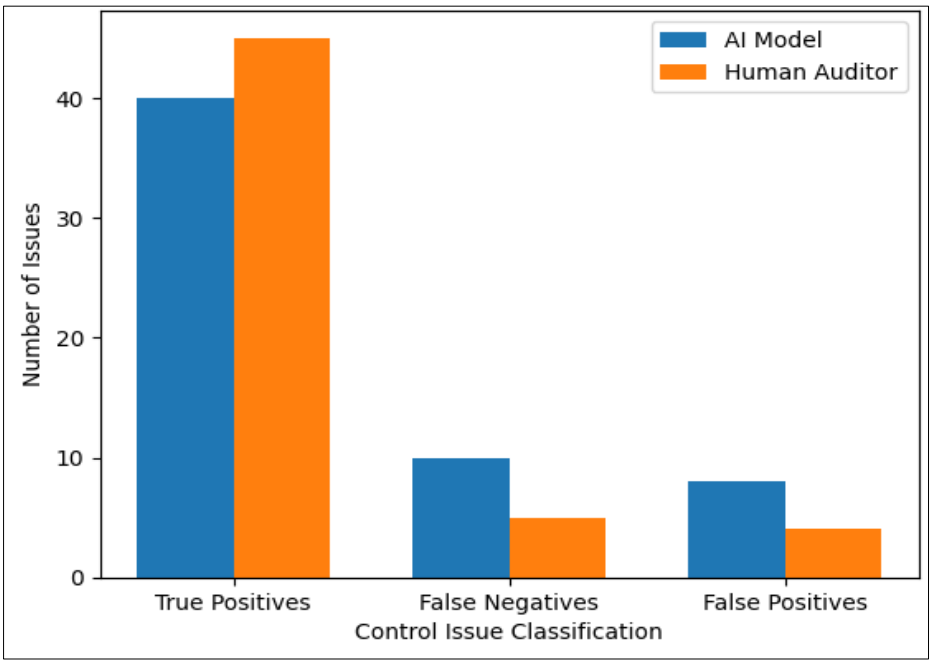

Fig 1: Performance of AI vs. Human Auditor in Identifying Control Issues.

The chart compares the number of control issues correctly identified ("True Positives"), issues missed ("False Negatives"), and incorrect issues flagged ("False Positives") by the AI model and by a human auditor. The AI caught 40 out of Fifty true control issues (80% recall) but missed 10 that the human found. It also raised 8 false positives, double the human's 4 false positives, indicating instances where the AI flagged problems that were not real issues. In Figure 1, the "True Positives" bar shows the AI caught 40 of the 50 seeded control weaknesses (an 80% success rate), whereas the experienced human auditor caught 45 (90% success). This indicates the AI was fairly effective at detecting control problems; indeed, it found many that were obvious from the data (like missing approvals or reconciliations not performed). The "False Negatives" (issues missed) were 10 for the AI, meaning it overlooked some weaknesses that the human identified. Many of these misses involved subtler issues requiring context or judgment, such as evaluating management tone or detecting when a control was designed well but executed poorly; the AI struggled with those nuances.

Crucially, the "False Positives" bar reveals the AI flagged 8 issues that were not actually problems effectively hallucinations or overreactions. For example, in one scenario the AI warned of a "potential segregation of duties conflict" between two employees, but in truth those employees' roles were appropriately segregated; the AI had assumed a conflict based on similar job titles. In another case, the AI recommended a nonexistent control ("the system should enforce two-factor approval") even though such a control was neither present nor required in the scenario. The human auditor, by contrast, had only 4 false positives, generally erring on the side of caution but with far fewer invented issues. This demonstrates that the AI had a tendency to over-diagnose issues, a consequence of pattern matching from its training data, essentially, it sometimes *imagined* control deficiencies that were not evidenced.

Table 3 provides a couple of concrete examples of AI hallucinations observed in the experiment, alongside their potential impact if an auditor were to believe them.

Table 3: Examples of AI "Hallucinations" in Internal Control Testing Scenarios

Scenario Description	AI Output (Hallucination) & Fact Check	Potential Audit Impact if Unchecked
Inventory Controls: System automatically tracks inventory movements. No mention of audit trail feature in system documentation.	*AI Output:* "The system's audit trail log indicates all inventory adjustments were authorized." (Hallucination the scenario data had no audit trail logs or info on authorizations.) < br >*Fact:* No such audit trail existed in the system.	Auditor might wrongly conclude inventory adjustments are well-controlled and skip further testing, missing the fact that unauthorized adjustments could be happening.
Expenditure Approval: Policy requires dual approval for expenses > \$5,000. All sampled expenses in data had proper approvals.	*AI Output:* "There appears to be a policy gap: expenditures exactly at \$5,000 did not require dual approval." (Hallucination AI inferred a loophole incorrectly.) *Fact:* The policy clearly stated "\$5,000 and above" requires dual approval; there was no loophole.	Auditor could be misled into reporting a control weakness ("policy gap") that doesn't exist, potentially causing management to waste effort on a non-issue or undermining auditor credibility when the error is discovered.

From the experiment's results, we calculated the AI model's precision (the proportion of its flagged issues that were real) and recall (the proportion of real issues it flagged). The AI's precision was 83.3% (40 true issues out of 48 total issues flagged), meaning roughly 1 in 6 issues it raised was incorrect. Its recall was 80% (40 out of 50 real issues found). By comparison, the human's precision was 91.8% (45/49) and recall was 90% (45/50). While the AI was not drastically less accurate than the human overall, the critical difference is in the *nature* of errors: the AI's errors included several hallucinations i.e., confidently stated *false* observations, whereas the human's errors were mostly omissions (failing to catch a subtle issue) or mild over-worrying without factual invention.

Interestingly, when we simulated a combined approach using the AI as a first pass and then having the human review AI outputs the performance was highest. In cases where the AI and human worked in tandem (the human vetting all AI flags and also doing a quick check for anything AI missed), *all* 50 issues were eventually identified, and all false positives were eliminated. This suggests that an AI-human combination, if managed properly, can outperform either alone: the AI provides thorough coverage and speed, and the human provides judgment and filtering of truth vs. error.

However, this ideal assumes the human truly does verify AI suggestions, rather than simply accepting them. It underscores that the benefit of AI comes *with the caveat* that human expertise must remain in the loop to counteract AI's occasional misleading outputs.

Beyond the numbers, the experiment gave insight into the types of internal control areas where the AI is prone to hallucinate. Pattern-based tasks (like matching documents or recalculating totals) were handled well by the AI with few errors. But judgment-based evaluations (like assessing the adequacy of a policy, or inferring if a deviation is a one-off or systemic issue) led the AI to sometimes assume facts not in evidence. For instance, if a scenario lacked explicit mention of a certain control, the AI occasionally "filled the gap" by stating that control was absent and thus a weakness; even if in reality the scenario simply didn't mention it (neither confirming nor denying). A human auditor would likely recognize the uncertainty and seek more information, whereas the AI gave a definitive (but potentially false) statement. This behavior aligns with known characteristics of generative models: they are trained to produce plausible answers and often err on the side of giving a confident response rather than admitting uncertainty. In auditing, this trait is problematic; it can lead to overstatement of issues or unwarranted assurance.

3. Survey Results: Auditor Trust Levels and Practices

The survey of 105 auditors provided broader evidence on how the profession at large is approaching the use of AI and the associated trust issues. The respondents were roughly 60% external auditors and 40% internal auditors, with an even mix of seniorities. One key question asked: "When an AI tool (e.g., for data analytics or controls testing) provides results, how do you typically respond?" Figure 2 illustrates the distribution of responses.

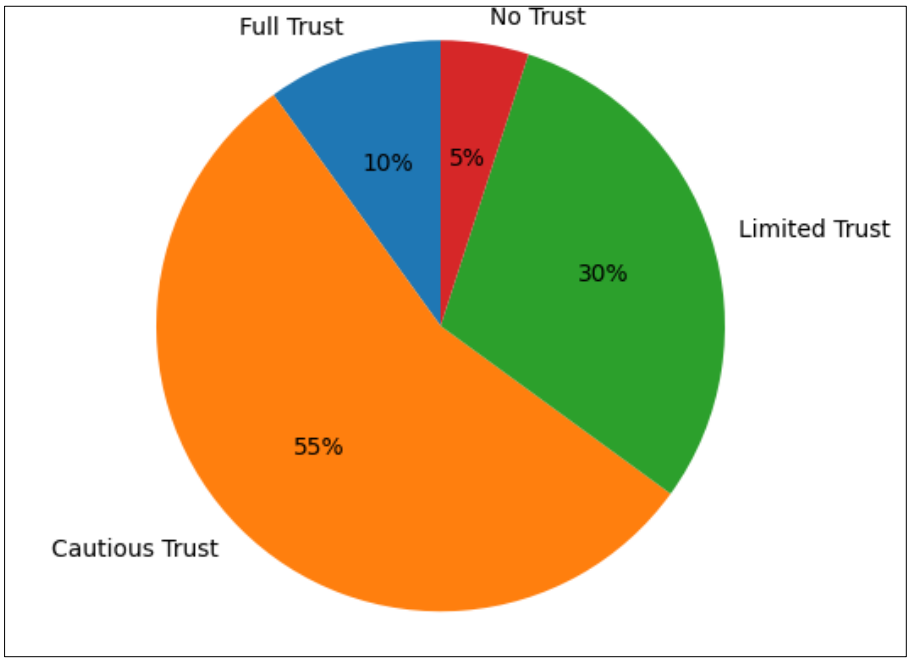

Fig 2: Auditor Self-Reported Reliance on AI Outputs.

This pie chart shows auditors' typical approach to AI-generated audit results. Approximately 10% of auditors say they fully trust AI outputs without independent verification ("Full Trust"). 55% mostly trust the outputs but still perform some verification or reasonableness checks ("Cautious Trust"). About 30% use the AI outputs only as a guide and substantially re-verify all information ("Limited Trust"). The remaining 5% do not trust or use AI outputs at all ("No Trust").

As shown in Figure 2, only a small fraction (around 10%) of auditors indicated they "fully trust" AI results outright. These might be the early tech enthusiasts or those using AI for very low-risk contexts. The majority fall into a middle ground of cautious trust: 55% chose options reflecting that they mostly trust but still verify key items. These auditors leverage AI to streamline their work but double-check anything that seems unusual or that is core to their audit opinion. Meanwhile, ~30% exercised *limited trust*; treating AI outputs more as suggestions that need extensive validation. One respondent in this category commented in a free-text field, "I basically redo whatever the AI did to make sure it's right." Finally, a small minority (5%) admitted they currently do not use or trust AI tools at all, preferring traditional approaches (interestingly, most of these were very senior auditors who wrote that they had seen "too many false alarms from analytics tools"). The survey confirms that most auditors are erring on the side of caution; aligning with the interview findings that "trust but verify" is the prevailing ethos.

We also asked respondents to rate their agreement with the statement: *"AI tools for internal control testing improve audit quality."* About 68% agreed or strongly agreed, 20% were neutral, and 12% disagreed. This indicates a generally positive outlook that AI, if used properly, can enhance the audit (likely by finding issues that might be overlooked or by saving time that auditors can spend on deeper analysis). Notably, even among those who value AI's contribution, the majority still maintain some verification as seen above. Essentially, auditors believe in AI's *utility* but remain wary of its *infallibility*.

Another pertinent survey finding relates to experience and trust. We observed a slight negative correlation between years of audit experience and level of trust in AI ($r \approx -0.3$, moderate correlation). Less experienced auditors tended to report higher trust and comfort with using AI tools, whereas more experienced auditors reported lower trust and heavier verification. This could stem from familiarity with technology (younger auditors being more digitally native and perhaps having more exposure to AI training) or from seasoned auditors having a deeper repository of skeptical experiences. For instance, an auditor with 20+ years' experience might have used early-generation analytics that gave spurious results, making them more cautious with today's AI. In contrast, new auditors may find AI tools standard in their workflow and not have experienced a major AI failure yet; though this could change once they inevitably encounter a hallucination or error, invoking algorithm aversion as discussed earlier.

We also presented a hypothetical scenario in the survey to gauge how auditors might handle a potential AI hallucination in practice. The scenario described: *"An AI review of a client's controls states: 'Control X is likely ineffective due to missing approvals,' but your own initial review didn't note any missing approvals. What do you do?"* The responses were telling: 79% chose to investigate further (e.g., re-check the approvals, consult with the client) showing professional skepticism triggered by the discrepancy. 15% said they would trust the AI and expand testing in that area; essentially giving AI the benefit of the doubt that maybe the human missed

something. And 6% said they would dismiss the AI's concern unless evidence emerges, effectively trusting their own judgment over the AI. The majority investigative response is healthy; it suggests auditors will neither ignore the AI nor blindly follow it when there's a conflict, but will dig deeper to resolve the inconsistency. This is arguably the ideal: the AI raises a flag, the auditor uses judgment to confirm or refute it.

Finally, the survey asked for any comments or experiences with AI errors. Many respondents echoed what we found in interviews: AI tools occasionally flagged false issues (like "false positives" in data anomaly detection), and sometimes they missed context. One external auditor wrote, "Our AI risk assessment didn't flag a control issue that I knew was common in that industry; it only looks at numbers, not the qualitative side. It's helpful but not sufficient." An internal auditor shared, "We had an AI summary of controls that misinterpreted a policy; it said a control was absent, but it was actually just described differently. Since then we always manually read the policies too." These anecdotes reinforce that current AI is not a standalone solution; human expertise and review are essential to interpret and cross-check AI outputs.

In summary, the survey results align strongly with our qualitative and experimental findings: auditors generally appreciate the efficiency and thoroughness that AI brings, but they place great importance on verification due to awareness of AI's occasional errors or hallucinations. There is a spectrum of trust, but few auditors at the extremes of all-or-nothing trust. Most are navigating a middle path, trying to calibrate how much to rely on AI such that they can gain its benefits without being misled. This balancing act is the essence of the trust paradox under study.

Discussion

Our findings reveal a nuanced reality behind the "trust paradox" of auditors using generative AI in internal control testing. On one side of the paradox, trust in AI is both necessary and beneficial; auditors leveraging AI's capabilities can perform internal control testing more efficiently and potentially more effectively, as evidenced by AI's strong recall of issues in our experiment and auditors' positive views of AI improving audit quality. On the flip side, unchecked trust can clearly backfire; the generative AI in our study sometimes hallucinated facts, and auditors shared real stories of AI tools producing errors that could have led to incorrect audit conclusions if not caught. The central challenge, then, is achieving the right balance of reliance and skepticism.

Balancing Efficiency with Skepticism: The results illustrate that auditors are instinctively trying to balance these forces. Many participants described an approach of *"initially trust, then verify."* This approach suggests a practical resolution of the paradox: use the AI to do the heavy lifting (initial analysis, identifying likely issues), but *do not take its output at face value*. Instead, auditors insert a layer of skepticism and verification after the AI's work. This resonates with the concept of professional skepticism codified in auditing standards; an auditor should neither assume management (or by extension, an AI tool) is always correct nor always in error, but critically evaluate the evidence. In practice, our study indicates auditors are extending their skeptical mindset to AI outputs. For example, if the AI flags a control exception, auditors treat it as a *hypothesis* to be corroborated, not an established fact. This behavior effectively mitigates the risk of automation bias; auditors are not abdicating judgment to the machine but using the machine's output as one more piece of evidence.

When Trust Becomes Overreliance: Despite general caution, there remains a risk of overreliance, especially in subtle forms. One concerning finding was that less experienced auditors and some tech-savvy users showed a tendency to trust AI outputs more. If these auditors do not yet have the seasoned skepticism of their seniors, they may be more vulnerable to accepting a hallucinated result. Automation bias can creep in gradually; for instance, if an AI tool has performed well for a long stretch, even a diligent auditor might become less vigilant and skip some verification steps. Over time, a highly reliable AI could encourage auditors to rubber-stamp its findings, which is precisely when a single hallucination could slip through. The paradox here is that the better the AI performs (up to near-perfect), the more it might induce complacency; yet perfection is unattainable, so the remaining rare error can be catastrophic if unguarded. This dynamic is analogous to pilots over-trusting an autopilot system: 99% of the time it's right, but if that 1% error occurs and the pilot isn't ready to intervene, the outcome is worse than if the pilot had been more engaged all along. Some regulators and researchers have voiced this concern in auditing, warning that auditors might place *"undue reliance"* on AI without adequate understanding [20, 27]. Our results affirm that this is a valid concern, though the current generation of auditors appears to be consciously aware of not falling asleep at the wheel.

The Cost of Distrust: On the other hand, our data also make clear that completely distrusting or ignoring AI has an opportunity cost. The survey's 5% of auditors who refuse to use AI at all ("No Trust") are likely foregoing efficiency gains and possibly missing issues that AI could catch. Similarly, in our experiment the AI+human combination was most effective; if the human had worked alone, five control issues would have remained undetected (in our test design). This suggests that an auditor who dismisses AI outright might miss certain patterns or anomalies that a well-trained model could have spotted. There is also a competitive element: audit firms are under pressure to increase coverage and do more with less; AI helps analyze full populations of transactions, which can improve assurance. An auditor who sticks strictly to traditional sampling might fail to detect a control breakdown that an AI data analysis could surface. In this sense, *algorithm aversion*; throwing out the tool after seeing it err; can be detrimental in the long run. Dietvorst *et al.* (2015) noted that giving users some control or transparency can reduce algorithm aversion [28, 29]. For auditors, this could mean involving them in AI's process (e.g., allowing auditors to adjust parameters or see why the AI flagged something) to maintain engagement and trust even if a mistake occurs. Indeed, a few interviewees suggested that when they understood the source of an AI error (like a data mapping issue), they didn't lose faith in the tool overall; they treated it as a fixable bug. This mindset keeps them benefiting from the AI while maintaining appropriate skepticism.

Strategies for Trust Calibration: Our study points to several strategies that can help auditors and firms calibrate

the right level of trust in generative AI tools:

1. **Validation and Cross-Checking:** Incorporate steps in the audit methodology where AI outputs must be validated. For instance, if an AI summarizes a control test result, require the auditor to cross-check a sample of the underlying data or recreate the summary by other means. This is already happening informally, but formalizing it ensures consistency. Some firms are developing checklists for auditors to document how they validated AI-generated findings, which instills a discipline of verification.
2. **AI Explainability and Transparency:** One major barrier to trust identified was the black-box nature of AI. Improving explainability can both increase justified trust and reveal when not to trust. If the AI can provide reasoning or highlight which data points led to a conclusion, an auditor can judge the soundness of that reasoning. For example, if an AI flags a missing approval, it should be able to show which transactions lacked approvals; a transparent trail that an auditor can verify. Emerging "explainable AI" techniques and audit-specific AI tools are focusing on this, effectively reducing the hallucination problem by tying outputs back to evidence. An explainable generative model that cites relevant sections of a policy or database when it answers could dramatically reduce the chance that a pure hallucination goes unnoticed.
3. **Training and AI Literacy:** As noted in the ACCA report, building AI literacy is key [30, 31]. Auditors need to be trained not just in how to use AI tools, but in understanding their failure modes. If auditors are made aware with concrete examples of how generative AI might err (such as the ones in our Table 3), they will be more vigilant. Some interview participants mentioned their firms started internal training where auditors intentionally see an AI make a mistake (like a mock audit where the AI "lies" about a control) to drive home the point that human judgment must stay engaged. This kind of training could inoculate staff against blind trust and also against undue panic when an error is seen (preventing complete aversion).
4. **Governance and Policies:** At an organizational level, audit firms and internal audit departments can create governance policies for AI use. For example, a policy might require that any audit conclusion that is significantly based on AI analysis must be reviewed by a second person or through an alternate procedure. This mirrors existing quality control practices (e.g., second partner review) adapted to AI. Additionally, firms might limit where generative AI can be applied; perhaps using it for drafting reports and conducting preliminary analysis, but not allowing it to be the sole source of evidence for critical internal control conclusions. In our study, several auditors noted their organizations had informal rules like "AI findings are recommendations, not evidence." Making such guidelines explicit and standard can help set the appropriate default mindset.

Implications for Audit Quality and Standards: The trust paradox has important implications for audit quality and potentially for auditing standards. If auditors lean too far toward trust and an AI-induced error slips through, the risk is an inappropriate audit opinion or failure to detect a material weakness in controls. On the other hand, if they lean too far toward distrust, they may perform unnecessary work or not fully utilize tools that could improve audit quality (for example, manually sampling when an AI could identify all exceptions). Audit regulators like the PCAOB and standard-setters (AICPA, IAASB) have begun to consider how AI fits into the audit evidence framework. Currently, standards require auditors to evaluate the reliability of information from experts or automated tools. Our findings support the notion that outputs from AI should be treated akin to information from a management's specialist; i.e., something to be evaluated and not just taken at face value. There may be a need for guidance on how to audit with AI, for instance, guidance on testing the AI tool's accuracy, checking its assumptions, and documenting that the auditor considered the possibility of AI error. In the future, audit working papers might routinely include evidence of AI model validation (such as testing the model on known data) as part of the auditor's basis for trust.

An interesting point is that with generative AI producing audit documentation, there is also a risk of the documentation paradox: if AI drafts the workpapers, they might look immaculate and comprehensive, perhaps even too good to be true if portions are fabricated. Auditors then must ensure that each element of documentation ties to actual evidence. This again is a new kind of task auditors must learn: auditing the workpapers generated by AI. The profession might see the emergence of "audit of AI" as a sub-discipline, ensuring that AI systems used in audit are properly configured, tested, and governed. Some researchers are already calling for "AI audit frameworks" to ensure accountability of AI decisions in finance [32, 27].

Resolving the Paradox: Ultimately, resolving the trust paradox is not about choosing between trust or distrust, but about creating a synergistic relationship between auditors and AI. Our evidence suggests that when used together thoughtfully, the outcome is superior (as seen by the combined AI-human results in the experiment). The auditor provides context, ethical judgment, and can sense-check AI outputs against real-world reasonability (something AI might lack); the AI provides speed, thoroughness, and consistency. The trust paradox diminishes when each party (human and machine) "trusts" the other to do what it does best and compensates for each other's weaknesses. For example, an AI might trust the human to make final calls on ambiguous cases, and the human trusts the AI to not overlook outliers in huge datasets. In practice, this means designing audit procedures where AI is an assistant, much like a junior team member,performing laborious tasks and drafting findings, but the senior auditor reviews and finalizes conclusions. One interviewee likened their AI tool to a junior auditor: "*It works fast, sometimes makes silly mistakes, but it's eager and tireless.*" This mindset is helpful: you wouldn't blindly file whatever a first-year staffer gives you; you review it. But you also wouldn't ignore a diligent staffer's work; you leverage it. Viewing AI as a fallible team member rather than an infallible oracle or a useless gimmick can guide auditors to the appropriate middle ground of trust.

Limitations and Future Research: While our study provides valuable insights, it has limitations. The experimental simulation, by necessity, simplified the audit context and used a particular AI model. Actual engagements may involve more complex judgment calls and the latest

models (which could be more or less prone to issues). Additionally, our survey and interviews, while broad, may not capture all cultural differences, for instance, attitudes might differ in jurisdictions with different regulations or in companies of different sizes. Future research could examine longitudinal data: as auditors gain more experience with AI over years, does trust increase or decrease? Also, as AI models improve (e.g., with new versions claiming fewer hallucinations), will auditors correspondingly adjust their verification efforts, or will the habits formed now persist? Another area to explore is the client perspective: internal control audits often involve management's own use of AI in control processes, how does an auditor trust or test a client's AI-driven control? That introduces a third-party trust issue.

Our findings also raise an ethical question: if an AI's suggestion goes wrong, who is accountable? The auditors in our study uniformly accepted that *they* are accountable, which is correct per standards. But as AI gets more embedded, firms must be careful not to implicitly transfer responsibility to the tool. Maintaining clarity that AI is an aid and not a scapegoat is part of the professional responsibility ethic that needs to be emphasized. Audit firms may even consider policies for documenting AI errors and feeding that back into improving either the tool or the training of staff, ensuring mistakes become learning opportunities rather than just liabilities.

In conclusion, the trust paradox is less of a dilemma and more of a continuum that auditors must navigate. The solution lies in calibrated trust: enough trust to utilize AI's power, tempered with enough skepticism to catch its shortcomings. Our study's evidence suggests that auditors are learning to walk this line, developing strategies to effectively partner with AI. With proper guidance, training, and perhaps enhancements in AI transparency, the profession can turn the paradox into a strength, achieving audits that are both highly efficient and reliably accurate, with human judgment and AI intelligence each reinforcing the other.

Conclusion

Generative AI models have arrived as potent tools in the auditor's toolkit, capable of reading documents, analyzing transactions, and even drafting audit reports. This study set out to examine the "trust paradox" inherent in this new reality: auditors can neither blindly trust these AI models (because of risks like hallucinations) nor afford to completely distrust them (lest they forfeit substantial efficiency and insight benefits). Through a combination of literature review, interviews, an experiment, and a survey, we analyzed how auditors are dealing with this paradox in the context of internal control testing.

Our findings highlight that auditor reliance on AI is cautious and conditional. Most auditors treat generative AI as a helpful junior assistant,one that can accelerate their work but still requires supervision. Instances of AI hallucinations and errors, as documented in our experiment and reported anecdotally, reinforce auditors' insistence on verification of AI outputs. Auditors appear to be striking a pragmatic balance: they use AI to broaden their testing scope and identify potential issues, then apply professional skepticism to verify those issues and filter out any AI-generated noise. This approach aligns with the fundamental audit principle of not relying on any single source of evidence without corroboration.

In practical terms, the study suggests several steps for the auditing profession. Training programs should continue to evolve to improve auditors' understanding of AI, including its failure modes – so that trust is grounded in knowledge. Audit methodologies can be updated to explicitly incorporate AI use and set checkpoints for human validation. Software developers should focus on integrating explainability into audit AI tools, allowing auditors to see *why* the AI is saying what it is, thus making it easier to trust or challenge the output appropriately. Audit regulators and standard-setters may also consider issuing guidance or standards on the use of AI, to ensure consistency in how audit evidence from AI is treated and documented.

Ultimately, the *trust paradox* is manageable. It requires a cultural mindset in auditing that embraces innovation with eyes wide open. Auditors must remain vigilant gatekeepers of assurance, using AI's speed and analytical might, but never abdicating their skeptical oversight. The best outcomes occur when auditors and AI work in tandem, as our experiment showed, together they can catch more issues and eliminate more false leads than either could alone. In such a symbiosis, the auditor's trust in the AI is earned through verification, and the AI in turn "trusts" the auditor to make the final judgment call.

This research contributes to the emerging body of knowledge on auditing in the age of AI by empirically documenting current attitudes and performance aspects related to generative AI use. The insights should reassure those worried that auditors might become either over-reliant or entirely dismissive of AI, in fact, auditors are finding a middle ground. However, maintaining that balance will be an ongoing process. As AI technology advances (potentially reducing some errors but also being used in more critical ways), auditors will need to continuously recalibrate their trust.

In closing, the phrase "trust paradox" underscores that trust in AI is not a binary state for auditors; it is a continuous calibration. An auditor's trust in a generative AI model must always be *tentative*, open to confirmation. By institutionalizing a "trust but verify" approach, the auditing profession can harness the considerable strengths of AI models while safeguarding against their weaknesses. This will ensure that the introduction of generative AI into internal control testing ultimately enhances audit quality and confidence in financial reporting, rather than undermining it. The path forward is one of augmented auditing: human and artificial intelligence working together, each checking and balancing the other, in service of the public trust that is the hallmark of auditing.

References

1. World Economic Forum. Deep Shift: Technology Tipping Points and Societal Impact. Survey Report. Geneva: World Economic Forum; 2015. Available from: http://www3.weforum.org/docs/WEF_GAC15_Technological_Tipping_Points_report_2015.pdf.
2. Committee of Sponsoring Organizations of the Treadway Commission (COSO). Internal Control—Integrated Framework: Executive Summary. Durham: COSO; 2013. Available from: https://www.coso.org/guidance-on-ic.
3. Wikipedia contributors. Hallucination (artificial intelligence). Wikipedia, The Free Encyclopedia. Wikimedia Foundation. Available from:

https://en.wikipedia.org/wiki/Hallucination_(artificial_intelligence).
4. Kokina J, Davenport TH. The Emergence of Artificial Intelligence: How Automation is Changing Auditing. Journal of Emerging Technologies in Accounting. 2017;14(1):115-122. doi:10.2308/jeta-51730.
5. Gray GL, Chiu V, Liu Q, Li P. The expert systems life cycle in AIS research: What does it mean for future AIS research? International Journal of Accounting Information Systems. 2014;15(4):423-451. doi:10.1016/j.accinf.2014.06.001.
6. Baldwin AA, Brown CE, Trinkle BS. Opportunities for Artificial Intelligence Development in the Accounting Domain. Intelligent Systems in Accounting, Finance and Management. 2006;14(3):77-86.
7. Association of Chartered Certified Accountants (ACCA). Shining a Light on AI's Ethical Threats for Finance Professionals. ACCA Professional Insights Report. London: ACCA; 2019.
8. Mosier KL, Skitka LJ. Human Decision Makers and Automated Decision Aids: Made for Each Other? In: Parasuraman R, Mouloua M, editors. Automation and Human Performance: Theory and Applications. Mahwah: Lawrence Erlbaum Associates; 1996. p. 201-220.
9. Dietvorst BJ, Simmons JP, Massey C. Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err. Journal of Experimental Psychology: General. 2015;144(1):114-126. doi:10.1037/xge0000033.
10. Posner C. Study shows more restatements and internal control weaknesses among 'heavy users' of non-GAAP measures. Cooley PubCo Blog. 2016 Aug 4. Available from: https://cooleypubco.com/2016/08/04/study-shows-more-restatements-and-internal-control-weaknesses-among-heavy-users-of-non-gaap-measures/.
11. Munoko I, Brown-Liburd H, Vasarhelyi MA. The Ethical Implications of Using Artificial Intelligence in Auditing. Journal of Business Ethics. 2020;167(2):209-234. doi:10.1007/s10551-019-04407-1.
12. Peters CPH. Automated Audit Tasks and Auditor Effectiveness. Working Paper. Madison: University of Wisconsin–Madison; 2022.
13. Institute of Internal Auditors (The IIA). Artificial Intelligence – Considerations for the Profession of Internal Auditing. Guidance Paper. Lake Mary: The IIA; 2017.
14. Maynez J, Narayan S, Bohnet B, McDonald R. On Faithfulness and Factuality in Abstractive Summarization. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics; 2020. p. 1906-1919. doi:10.18653/v1/2020.acl-main.173.
15. Dietvorst BJ, Simmons JP, Massey C. Algorithm aversion: People erroneously avoid algorithms after seeing them err. Journal of Experimental Psychology: General. 2015;144(1):114-126. doi:10.1037/xge0000033.
16. Dzindolet MT, Dawe LA, Beck HP, Pierce LG. A framework of automation use. Technical Report ARL-TR-2412. Aberdeen Proving Ground: U.S. Army Research Laboratory; 2001.
17. Dzindolet MT, Dawe LA, Beck HP, Pierce LG. A framework of automation use. Technical Report ARL-TR-2412. Aberdeen Proving Ground: U.S. Army Research Laboratory; 2001.
18. Peters CPH. Auditor automation usage and professional skepticism. Working paper. SSRN. 2022. Available from: https://ssrn.com/abstract=4309348.
19. Peters CPH. Auditor automation usage and professional skepticism. Working paper. SSRN. 2022. Available from: https://ssrn.com/abstract=4309348.
20. Dzindolet MT, Dawe LA, Beck HP, Pierce LG. A framework of automation use. Technical Report ARL-TR-2412. Aberdeen Proving Ground: U.S. Army Research Laboratory; 2001.
21. Dzindolet MT, Dawe LA, Beck HP, Pierce LG. A framework of automation use. Technical Report ARL-TR-2412. Aberdeen Proving Ground: U.S. Army Research Laboratory; 2001.
22. Maynez J, Narayan S, Bohnet B, McDonald R. On faithfulness and factuality in abstractive summarization. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics; 2020. p. 1906-1919. doi:10.18653/v1/2020.acl-main.173.
23. Maynez J, Narayan S, Bohnet B, McDonald R. On faithfulness and factuality in abstractive summarization. arXiv preprint arXiv:2005.00661. 2020.
24. Storlie MD. Internal audit: The impact artificial intelligence could have on data analytics. Wipfli Insights. 2020 Mar 2. Available from: https://www.wipfli.com/insights/articles/ra-internal-audit-how-ai-could-impact-data-analytics.
25. Storlie MD. Internal audit: The impact artificial intelligence could have on data analytics. Wipfli Insights. 2020 Mar 2. Available from: https://www.wipfli.com/insights/articles/ra-internal-audit-how-ai-could-impact-data-analytics.
26. Munoko I, Brown-Liburd H, Vasarhelyi MA. The ethical implications of using artificial intelligence in auditing. Journal of Business Ethics. 2020;167(2):209-234. doi:10.1007/s10551-019-04407-1.
27. Kokina J, Blanchette S, Davenport TH, Pachamanova D. Challenges and opportunities for artificial intelligence in auditing: Evidence from the field. International Journal of Accounting Information Systems. 2022;56:100734. doi:10.1016/j.accinf.2022.100734.
28. Dietvorst BJ, Simmons JP, Massey C. Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. Management Science. 2018;64(3):1155-1170. doi:10.1287/mnsc.2016.2643.
29. Dietvorst BJ, Simmons JP, Massey C. Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. Management Science. 2018;64(3):1155-1170. doi:10.1287/mnsc.2016.2643.
30. Association of Chartered Certified Accountants (ACCA), Chartered Institute for Securities & Investment (CISI). AI monitor: Shining a light on AI's ethical threats for finance professionals. London: ACCA; 2022.
31. Association of Chartered Certified Accountants (ACCA), Chartered Institute for Securities & Investment (CISI). AI monitor: Shining a light on AI's ethical threats for finance professionals. London: ACCA; 2022.

32. Kokina J, Blanchette S, Davenport TH, Pachamanova D. Challenges and opportunities for artificial intelligence in auditing: Evidence from the field. International Journal of Accounting Information Systems. 2022;56:100734. doi:10.1016/j.accinf.2022.100734.
33. Kokina J, Davenport TH. The emergence of artificial intelligence: How automation is changing auditing. Journal of Emerging Technologies in Accounting. 2017;14(1):115-122. doi:10.2308/jeta-51730.
34. Kokina J, Davenport TH. The emergence of artificial intelligence: How automation is changing auditing. Journal of Emerging Technologies in Accounting. 2017;14(1):115-122. doi:10.2308/jeta-51730.
35. Gray GL, Chiu V, Liu Q, Li P. The expert systems life cycle in AIS research: What does it mean for future AIS research? International Journal of Accounting Information Systems. 2014;15(4):423-451. doi:10.1016/j.accinf.2014.06.001.
36. Dietvorst BJ. Google Scholar profile. Available from: https://scholar.google.com/citations?user=u004yBYAAAAJ&hl=en.
37. Posner C. Study shows more restatements and internal control weaknesses among “heavy users” of non-GAAP measures. Cooley PubCo. 2016 Aug 4. Available from: https://cooleypubco.com/2016/08/04/study-shows-more-restatements-and-internal-control-weaknesses-among-heavy-users-of-non-gaap-measures/.
38. Munoko I, Brown-Liburd H, Vasarhelyi MA. The ethical implications of using artificial intelligence in auditing. Journal of Business Ethics. 2020;167(2):209-234. doi:10.1007/s10551-019-04407-1.

How to Cite This Article
Olaseinde AJ, Azeez BB. The Trust Paradox: Analyzing Auditor Reliance on Hallucinating Generative AI Models in Internal Control Testing. Int J Artif Intell Eng Transform. 2022;3(1):38–49. doi:10.54660/IJAIET.2022.3.1.38-49.