Effort, Risk, and Retention: A Behavioral Model of Human Auditors and Their AI Rivals

Sangphill Kim
University of Massachusetts, Lowell
Taewoo Kim*
California State University, San Bernardino

ABSTRACT

This paper presents a behavioral model comparing human auditors and AI rivals in a competitive audit market. While many assume AI is superior, explanations are often vague. We show that AI's advantage emerges directly from human constraints: risk aversion and effort aversion. Assuming fixed audit fees, we model audit quality as a function of effort and perceived risk. In a multi-period setting, we derive a critical retention threshold that governs when firms prefer humans despite lower audit quality. Our model explains this misalignment and proposes policy reforms to align retention incentives, assign legal responsibility, and preserve audit quality in an AI-integrated future.

Keywords: Audit quality, Artificial intelligence, Risk aversion, Effort aversion, Client retention, Audit policy

JEL: M42, C61, D91, L84, O33

*Contact author: taewoo.kim@csusb.edu

I. INTRODUCTION

As artificial intelligence (AI) continues to reshape professional services, the field of auditing faces a pivotal shift. AI systems now assist in fraud detection, anomaly identification, and data classification—functions that once required seasoned human judgment (Appelbaum, Kogan, & Vasarhelyi, 2017; Brown-Liburd, Issa, & Lombardi, 2015). It is widely assumed, often without formal justification, that AI will soon outperform human auditors in both accuracy and efficiency. However, beneath this growing consensus lies a surprising gap: while empirical studies document the spread of AI in auditing, few have rigorously modeled why AI should outperform human auditors—or under what conditions that advantage may fail to materialize.

This paper addresses that gap by offering a structured behavioral comparison between human auditors and their AI rivals.[1] We do not ask whether AI is better; we ask why—and more importantly, under what assumptions. Most claims about AI superiority rely on its computational power or data capacity. Our model demonstrates that AI's advantage is far simpler and more fundamental: human auditors are risk-averse and effort-averse, while AI systems are risk-neutral and effort-neutral. These behavioral differences, when combined with fixed audit fees in a competitive market (Simunic, 1980; Hay, Knechel, & Wong, 2006), lead directly to divergences in audit quality.

We model audit quality as a function of auditor effort, subject to diminishing returns and a regulatory threshold. In a single-period setting, the result is clear: AI auditors achieve higher audit quality with lower internal friction, since they do not experience disutility from effort nor fear the consequences of failure (Antle & Nalebuff, 1991; DeAngelo, 1981). However, most real-world audit engagements are multi-period, where client retention plays a central economic role (Ashton & Ashton, 1988; Chang & Hwang, 2003). We therefore extend the model to a dynamic setting in which future audit fees depend on the probability of being re-hired.

In this extended framework, human auditors gain a crucial edge—not through technical superiority, but through relationship stickiness: the tendency of clients to favor familiar auditors based on interpersonal trust, inertia, or legacy ties (DeAngelo, 1981; Rennie, Kopp, & Lemon, 2010). AI systems, despite producing objectively superior audit output, may face retention friction due to perceived coldness, lack of explainability, or unclear accountability (Lacmanovic and Skare, 2025).

This asymmetry between performance and retention has major implications. It raises fundamental questions about auditor selection, audit quality, and market efficiency. Should the audit profession optimize for technical quality or for relational continuity? Should policy prioritize innovation or stability? Should regulators intervene to correct the disconnect between output and reward?

To address these questions, we propose a set of policy responses grounded in the model. These include: hybrid audit frameworks in which AI handles analysis while humans retain legal and interpretive responsibility; mandatory retention disclosures to highlight the basis of client loyalty; and accountability reforms that clarify liability in AI-driven audits. Our contribution is to offer a formal behavioral model that explains the superiority of AI auditors under economic constraints, shows how that superiority may be undercut by human relationships, and offers a path forward for aligning audit quality with long-term governance.

---

[1] Consistent with the definition provided by PWC (2017), the AI auditors in this paper are understood as *Autonomous* AI. This classification is characterized by "automating decision making processes without human intervention" and includes "AI systems that can adapt to different situations and can act autonomously without human assistance.

One of the central contributions of this paper is the formal derivation and interpretation of a critical retention threshold R* which defines the minimum level of client loyalty required to make human auditors economically viable in the presence of AI. We show how this threshold depends on the structure of costs, fees, and effort, and how exceeding it may result in the rational selection of lower-quality auditors. We further quantify the resulting audit quality loss ΔQ and explain how both expressions can be used by policymakers to design retention caps, disclosure regimes, and accountability frameworks that preserve audit integrity in a market increasingly shaped by automation.

II. LITERATURE REVIEW

This section reviews five key strands of literature that form the foundation of our behavioral model: (1) the competitive structure of the audit market and audit fee rigidity, (2) risk and effort aversion among human auditors, (3) the economics and psychology of client retention, (4) the rise of AI in auditing and the barriers to its adoption, and (5) the gap in current modeling that motivates our contribution. Each section grounds a major component of our theoretical framework and demonstrates how this paper fits within — and moves beyond — the existing literature.

2.1 The Competitive Audit Market and Fee Rigidity

Much of the analytical and empirical auditing literature assumes that audit firms operate in a competitive market where pricing power is limited and audit fees are determined by external conditions (Simunic, 1980; Hay, Knechel, & Wong, 2006). Simunic's foundational model demonstrated that when reputation and regulatory oversight are strong, audit pricing converges toward marginal cost, making auditors effective price takers. This assumption has been repeatedly validated in empirical studies, including meta-analyses by Hay et al. (2006) and Francis (2011), who found that audit fees are more responsive to client risk and complexity than to firm-specific bargaining power.

Knechel and Willekens (2006) expand on this by showing that audit demand is shaped more by risk and governance than by pricing variation, suggesting that audit markets behave competitively even in oligopolistic settings. DeFond and Zhang (2014) similarly note that price sensitivity is often constrained by benchmark comparisons and regulatory expectation, especially for public companies. The result is what some scholars describe as "quasi-fixed" pricing — audit firms cannot charge significantly more for additional effort unless mandated by regulation or risk signals (Ghosh & Siriviriyakul, 2018; Yang, Lee, Lim, & Yi, 2021).

We adopt this assumption of fixed audit fees as a modeling simplification — not because pricing variation is impossible, but because our interest is in effort and utility, not price-setting strategy. Holding fees constant allows us to isolate the effect of effort and behavioral preference differences on audit quality.

2.2 Human Auditor Constraints: Risk and Effort Aversion

The behavioral economics of auditing has long emphasized that human auditors do not act as frictionless optimizers. Instead, they are constrained by risk aversion, effort sensitivity, and subjective perceptions of failure. Antle and Nalebuff (1991) were among the first to incorporate auditor conservatism into utility-maximizing models, showing that perceived litigation risk and personal liability drive underinvestment in marginal effort.

DeAngelo (1981) laid the groundwork for auditor behavior under asymmetric penalties, noting that auditor independence is difficult to maintain when firms internalize the reputational cost of audit failure. This creates a fear-driven ceiling on how far human auditors will push effort, especially when audit failure is uncertain but personally costly.

Effort aversion has also been explored directly. Annelin (2022) show that audit effort increases audit quality, but auditors often underperform relative to theoretical optimums, particularly when constrained by engagement budgets or partner-level fatigue. Bell, Landsman, and Shackelford (2001) argue that audit effort is shaped not only by risk expectations but also by real resource constraints — time, capacity, and attention — leading to rational under-effort when audit risk is perceived as manageable or dispersed.

Francis and Yu (2009) find that auditor size and staffing capacity correlate positively with audit quality, further suggesting that effort availability — not just knowledge — drives quality outcomes. Together, these findings support our assumption that human auditors are both risk-averse and effort-averse — and that these features meaningfully impact their decision-making under fixed-fee conditions.

2.3 Client Retention, Relationship Stickiness, and Independence Risks

One of the most persistent challenges in audit regulation is the stickiness of client relationships. While continuity may improve audit familiarity and efficiency (Ashton & Ashton, 1988), it also undermines auditor independence — especially when client loyalty becomes a function of interpersonal warmth rather than audit quality (DeAngelo, 1981). Chang and Hwang (2003) empirically show that auditors with strong retention incentives are more likely to accept aggressive financial reporting, particularly when client business risk is low. Singer and Zhang (2018) find that long-standing auditor-client relationships are associated with less timely discovery and correction of misstatement.

These findings have prompted policy responses. In the U.S., PCAOB Rule 203 mandates lead audit partner rotation every five years, while the European Union has imposed mandatory firm rotation after ten years for public interest entities. These rules are based on the assumption that relationship longevity may erode skepticism, especially when familiarity is unaccompanied by rising audit quality.

In our model, we formalize this idea using a retention function that depends both on objective audit quality and a human relationship parameter (H). While AI systems can match or exceed quality levels, they cannot build interpersonal capital — giving human auditors a long-run advantage in client retention even when they underperform on audit output.

2.4 AI in Auditing: Promise, Perception, and Policy Friction

There is broad agreement that AI has the potential to transform auditing. Appelbaum et al. (2017) outline how AI can automate anomaly detection, risk scoring, and substantive testing, while Brown-Liburd et al. (2015) show that machine learning tools outperform human auditors in complex judgment settings. Fedyk, Hodson, Khimich, & Fedyk (2022) document AI is associated with improved audit quality and reduced fees and is predicted ultimately displace human auditors.

Yet enthusiasm is tempered by limitations. Lehner et al. (2022) identify five ethical hurdles in AI-driven auditing decisions—objectivity, privacy, transparency, accountability, and trustworthiness—highlighting the importance of skilled AI application and governance while recognizing its constraints. Law and Shen (2025) also suggest that the adoption of AI transforms the role of auditors, requiring a different skill set and leading to enhanced audit quality, rather than displacing them. This mismatch discourages full automation and may delay adoption even where quality gains are provable.

Alles (2022) goes further, calling for a new governance model for AI-enabled audits — one that includes oversight protocols, ethical guidance, and disclosure rules to manage client perceptions and stakeholder accountability. These perspectives support our policy recommendations in later sections and confirm that technology alone cannot resolve the behavioral frictions embedded in audit decision-making.

2.5 Gaps in the Literature and Contribution of This Paper

The existing literature provides robust empirical and theoretical insights into audit pricing, effort, retention, and AI adoption. But no existing model combines these insights into a behavioral structure that allows formal comparison between human auditors and AI systems in a competitive market. Studies show that AI improves efficiency; others show that human biases exist. But few connect those findings in a way that quantifies the trade-offs, or that explores how market outcomes may favor underperformers due to perception-based retention.

Our model bridges that gap. By assuming fixed audit fees and defining audit quality as a function of risk-weighted effort, we show how AI achieves superior audit performance. By layering in a multi-period retention function, we show how humans retain clients through relational stickiness. And by connecting these outcomes to policy, we offer a pathway for improving audit governance in the presence of automation.

To address this gap, we introduce a formal behavioral model comparing human auditors and AI systems in a competitive audit market. The model incorporates three key components: (1) fixed audit fees reflecting price-taking behavior, (2) differences in risk and effort aversion between human and AI agents, and (3) a retention function that captures both audit quality and interpersonal relationships. In the single-period case, we demonstrate how AI achieves higher audit quality due to structural advantages. In the multi-period case, we examine how relationship stickiness allows human auditors to retain clients despite lower quality. The next section presents this framework in full.

III. MODEL OVERVIEW AND STRATEGY

This section presents the formal framework we use to compare audit quality across two types of auditors: a human auditor, who is risk-averse and effort-averse, and an AI auditor, who is risk-neutral and effort-neutral. The model unfolds in two steps. We begin with a single-period setting, where there is no client retention and each auditor's goal is to maximize utility within a single engagement. In this environment, we isolate the impact of behavioral traits on audit quality, holding the audit fee constant.

We assume a perfectly competitive audit market, consistent with Simunic (1980) and Hay, Knechel, and Wong (2006), in which audit fees are fixed and externally determined. Auditors cannot differentiate themselves on price; they compete purely on quality and efficiency. This assumption allows us to focus entirely on the auditor's effort decision, unconfounded by pricing strategies or reputational spillovers.

Audit quality is modeled as an increasing, concave function of effort, with a minimum threshold below which an audit is considered a failure. Both auditor types choose effort levels to maximize utility, subject to this quality constraint. In this single-period environment, we show that AI systems achieve higher audit quality not because they are "smarter," but because they lack the internal frictions that constrain human behavior. The human auditor, facing real psychological costs—litigation fear, fatigue, and limited attention—chooses a suboptimal effort level. AI, by contrast, allocates effort until the quality threshold is met, and does so with lower marginal cost. This setup provides a clean, controlled environment in which we can demonstrate the first key insight of the paper: the superiority of AI auditors arises naturally from differences in behavioral structure, even when all other conditions are held equal.

In the next subsection, we formally define the utility functions, cost structures, and audit quality function that govern the single-period model. Full derivations follow, leading to a clear comparison between optimal effort, utility, and audit quality across the two agent types.

3.1 The Single-Period Effort Model

We begin with a single period setting in which the audit firm is hired for one engagement, receives a fixed audit fee F, and chooses effort eee to maximize utility. This setup allows us to isolate the behavioral effects of effort and risk preferences under the assumption of a perfectly competitive audit market, where prices are exogenous and reputation effects are absent (Simunic, 1980; Hay, Knechel, & Wong, 2006). In this framework, we compare two types of auditors: a human auditor, who is risk-averse and effort-averse, and an AI auditor, who is risk-neutral and effort-neutral.

Audit quality Q(e) is modeled as a concave function of effort:

$$Q(e) = q \cdot \ln(1+e)$$

where $q > 0$ is a productivity parameter and $e \geq 0$ is the auditor's chosen effort. A minimum audit quality threshold Q_{min} must be met to avoid audit failure. If quality falls below this threshold, the audit firm incurs a penalty P, representing litigation risk, regulatory action, or reputational damage.

3.1.1. Human Auditor

The human auditor's utility reflects two behavioral frictions: the cost of effort and the perceived risk of failure. Effort is costly in a convex fashion, and failure is penalized based on the auditor's subjective perception of risk. The utility function is:

$$U_H = F - C_H \cdot e^2 - \rho_H \cdot \pi_H(e) \cdot P$$

where $C_H > 0$ represents the marginal cost of effort, $\rho_H \in (0,1)$ captures risk aversion, and $\pi^H(e)$ is the perceived probability of audit failure. We assume the human perceives failure probability as:

$$\pi_H(e) = e^{-k_H e}$$

with $k_H > 0$ reflecting how quickly perceived risk declines with effort. Substituting, we get:

$$U_H = F - C_H \cdot e^2 - \rho_H \cdot e^{-k_H e} \cdot P$$

The first-order condition (FOC) for utility maximization is:

$$\frac{dU_H}{de} = -2C_H e + \rho_H \cdot k_H \cdot e^{-k_H e} \cdot P = 0$$

The optimal effort e_H^* can be solved numerically, and the resulting audit quality is:

$$Q_H^* = q \cdot \ln(1 + e_H^*)$$

3.1.2. AI Auditor

The AI auditor is effort-neutral and risk-neutral. It does not experience disutility from effort, nor does it internalize any penalty from failure. Its utility is simply:

$$U_{AI} = F - C_{AI} \cdot e$$

with $C_{AI} \ll C_H$ reflecting a flat or near-zero marginal cost of computational effort. The AI chooses the minimum effort that satisfies the audit quality threshold:

$$q \cdot \ln(1+e) \geq Q_{min} \Rightarrow e_{AI}^{min} = \exp\left(\frac{Q_{min}}{q}\right) - 1$$

Then:

$$U_{AI} = F - C_{AI}\, e_{AI}^{min}.$$

and:

$$Q_{AI}^* = q - \ln(1 + e_{AI}^{min}) = Q_{min}$$

3.1.3. Comparison

Although both agents aim to meet the quality threshold, their internal constraints differ. The AI auditor chooses the minimal necessary effort to meet Q_{min} and does so at lower cost. The human auditor may fall short, over-exert due to risk aversion, or under-exert due to effort sensitivity. Even when the human achieves threshold quality, it does so with lower efficiency.

The model implies a clear inequality:

$$\Delta Q^* = Q_{AI}^* - Q_H^* > 0$$

This result follows from the structural differences: humans face psychological and economic frictions; AI does not. Even under fixed fees, AI produces higher utility and audit quality, demonstrating its superiority in a single-period setting.

3.2.1 Multi-Period Utility Framework

To capture the role of client retention in auditor selection, we extend the model into a multi-period setting. Unlike the single-period environment, where audit quality and cost efficiency determine selection, multi-period engagements introduce a dynamic element: the potential for repeat business. The auditor's value to the firm is no longer based solely on current-period performance but also on the probability of being retained in the future.

We model the engagement over two periods, with the audit fee F held constant in both. The firm evaluates the net utility of each auditor type by combining the current-period payoff with the discounted value of expected future payoffs, conditional on retention. Let $\beta \in (0,1)$ represent the discount factor.

The firm's utility from hiring a human auditor is:

$$U_H^{firm} = F - C_H e_H^2 + \beta \cdot R \cdot (F - C_H e_H^2)$$

where: $C_H e_H^2$ is showing the convex cost of human effort and R is the probability that the human auditor is retained for a second period

For an AI auditor, who is not retained due to lack of relationship stickiness, utility is strictly single period:

$$U_{AI}^{firm} = F - C_{AI} . e_{AI}$$

This distinction sets the foundation for the comparative analysis that follows. In what follows, we formalize the utility structure for each auditor type, derive the firm's selection condition, and show how retention alone can drive the choice of a lower-quality auditor.

3.2.2 AI Auditor Utility Function

We begin with the AI auditor, whose behavior is analytically simpler due to two core assumptions: risk neutrality and effort neutrality. The AI system does not internalize penalties associated with audit failure, nor does it experience disutility from exerting effort—aside from a linear cost tied to energy, computation, or system processing.

As in the single-period model, audit quality is a concave function of effort, given by:

$$Q(e) = q \ln (1+e)$$

The AI auditor selects the minimum level of effort needed to satisfy the regulatory audit quality threshold Q_{min}. This results in:

$$Q_{AI}^* = Q_{min} = e_{AI}^* = \exp\left(\frac{Q_{min}}{q} \right)$$

With cost of effort modeled linearly as C_{AI} e, the AI auditor's total cost becomes:

$$C_{AI} = C_{AI} . e_{AI}^*$$

And the firm's utility from hiring the AI auditor is:

$$U_{AI}^{firm} = F - C_{AI} e_{AI}^*$$

This outcome defines the AI benchmark: minimum necessary effort, maximum efficiency, and no retention benefit. The AI does not gain or lose from client relationships, making its expected utility strictly tied to single-period performance.

3.2.3 Human Auditor Utility Function

Unlike the AI system, the human auditor faces behavioral and economic frictions that alter their effort decision. We assume the human auditor is risk-averse and effort-averse. Effort comes with increasing marginal disutility, modeled as a convex cost function. In addition, the auditor internalizes the perceived risk of audit failure, which generates a potential penalty P if quality falls below the threshold Q_{min}.

Audit quality remains defined by:

$$Q(e) = q\cdot \ln(1 + e)$$

The perceived probability of audit failure is modeled as an exponentially declining function of effort:

$$\pi_H(e) = e^{-k_H e}$$

where $k_H > 0$ captures the auditor's sensitivity to effort as a buffer against risk.

The human auditor's utility over two periods incorporates: (1) Current Perid Payoff, (2) Convex Effort Cost, (3) Expected Penalty from Audit Failure and (4) Retention-Adjusted Future Utility.

Let $\rho_H \in (0,1)$ represent the auditor's risk aversion. Then, the firm's utility from hiring a human auditor is:

$$U_H^{firm} = F - C_H\, e_H^2 - \rho_H \,.\, \pi_H(e_H)\cdot P + \beta\cdot R\cdot (F - C_H\, e_H^2)$$

Here:

- $C_H\, e_H^2$: convex cost of human effort
- $\rho_H\cdot \pi_H(e_H)\cdot P$: expected penalty internalized by the auditor
- $\beta\cdot R\cdot(\cdot)$: discounted value of retention, where R is the retention probability

Compared to the AI auditor, the human must balance present risk, effort cost, and the possibility of future engagement. The optimal effort level e_h^* is chosen to maximize this utility, often requiring numerical solution. The resulting audit quality is:

$$Q_H^* = q \,.\, \ln(1 + e_H^*)$$

In this setting, the human auditor's strength lies not in efficiency, but in relational advantage the ability to be rehired due to trust, familiarity, or institutional history, as captured by R. The implications of this retention channel are explored next.

3.2.4 Retention as a Strategic Advantage

Retention plays a critical role in determining auditor selection in the multi-period setting. While AI auditors may deliver higher technical quality, their lack of interpersonal presence, explainability, or familiarity means they are unlikely to be retained. In contrast, human auditors benefit from relationship stickiness—the tendency of clients to prefer familiar service providers even in the presence of marginally better alternatives.

In our model, this advantage is captured by the parameter R, representing the probability that a human auditor is rehired in the second period. For AI auditors, we assume R = 0, either because they are replaced regularly by design, or because client sentiment and regulatory uncertainty limit their retention.

The firm's decision to hire a human auditor therefore becomes a function of two variables: *The Present Value of Retention-based Future Fees* and *The Cost and Quality* Disadvantage relative to AI.

From Section 3.2.3, we recall that the firm's utility from hiring a human auditor is:

$$U_H^{firm} = F - C_H\, e_H^2 - \rho_H \,.\, \pi_H(e_H)\cdot P + \beta\cdot R\cdot (F - C_H\, e_H^2)$$

The marginal value of retention is:

$$\frac{dU_H^{firm}}{dR} = \beta\cdot (F - C_H\, e_H^2)$$

This derivative shows that even moderate retention probabilities can meaningfully improve the firm's total utility from hiring a human auditor—particularly when audit fees are high or effort costs are relatively low. However, this selection mechanism is not based on audit quality**.** It is based on the economic value of continuity, and therefore, may lead to the selection of a lower-quality audit agent.

This is the central asymmetry revealed by our model: retention creates a structural channel through which lower-quality auditors can be chosen rationally, not because they outperform, but because they return. The next section formalizes this trade-off and introduces the critical threshold R^*, above which human auditors become preferred despite producing lower audit quality.

3.2.5 The Critical Retention Threshold and Audit Quality Loss

In the multi-period setting, human auditors benefit from relationship stickiness, allowing them to retain clients even when their audit quality is lower than that of AI auditors. While this dynamic is realistic, it raises a question of both economic and regulatory significance: under what conditions will a firm prefer to hire a lower-quality human auditor over a higher-quality AI system, solely because of retention incentives?

To answer this, we define the firm's objective as maximizing total expected net utility across two periods. For the AI auditor, the firm receives only one-period net utility, since AI lacks client stickiness and is not retained. For the human auditor, the firm benefits from the possibility of continued business, which is driven by the retention probability R. We assume audit fees F are fixed and that effort levels e_H and e_{AI} produce corresponding costs and audit quality.

Let:

- $C_H\, e_H^2$ cost of human audit effort
- $C_{AI} .e_{AI}$ cost of AI audit effort

- β : discount factor
- R: human auditor's retention probability

Then the firm's net utility from each auditor type is:

$$U_{AI}^{firm} = F - C_{AI} \,.\, e_{AI}$$

$$U_{H}^{firm} = F - C_H\, e_H^2 + \beta \cdot R \cdot (F - C_H\, e_H^2)$$

The firm is indifferent between the two when $U_{AI}^{firm} = U_{H}^{firm}$ which yields the critical retention rate R^* at which the human auditor becomes competitive:

$$R^* = \frac{C_H\, e_H^2 - C_{AI} \,.e_{AI}}{\beta \cdot (F - C_H\, e_H^2)}$$

This result defines the minimum level of relationship stickiness required for the human auditor to be economically viable. Below this threshold, AI dominates; above it, the human may be retained even at a cost to audit quality. This leads directly to the next key result: how much audit quality is sacrificed when the human auditor is chosen due to high retention**.** Define the quality gap:

$$\Delta.Q = Q_{AI}^* - Q_H^* = Q_{min} - q \,.\, \ln(1 + e_H^*)$$

This gap represents the audit quality penalty society incurs when a human auditor is selected for retention-driven reasons. Although firms may prefer human auditors for rational financial reasons—namely, future revenue—the choice comes at the cost of reduced assurance quality. This is especially problematic when market-wide retention practices lead to systemic reductions in audit performance, undermining public trust. The tension is sharpened by the fact that the firm's optimization problem and society's objective are not aligned. The firm seeks to maximize economic utility, including the value of repeat engagements. Society, however, prioritizes audit quality and independence—particularly in public markets where information asymmetry and risk are higher.

We formalize this divergence in the next section by defining the critical retention threshold R*, the point beyond which the human auditor becomes economically favorable to the firm, despite delivering lower-quality audits. We then quantify the audit quality loss and interpret its policy implications.

3.2.6 Interpreting the Critical Retention Threshold

We now turn to the interpretation of the critical retention threshold R^*, which defines the minimum level of client stickiness required for a human auditor to be chosen over an AI auditor in a two-period model. Recall that:

$$R^* = \frac{C_H\, e_H^2 - C_{AI} \,.e_{AI}}{\beta \cdot (F - C_H\, e_H^2)}$$

This expression reveals that the firm will prefer to hire a human auditor only if the expected value of retention outweighs the cost disadvantage associated with lower efficiency**.** That trade-off is captured entirely by the behavior of R*, which in turn depends on several structural parameters: the cost of effort for humans and AI, the effort levels required to meet quality thresholds, the audit fee F, and the discount factor β.

To better understand how these variables influence the retention threshold, we analyze the partial derivatives of R^* with respect to key inputs.

Partial Derivative Analysis

Let's explore the sign and interpretation of each:

(1) Cost of Human Effort: C_H

$$\frac{\partial R^*}{\partial C_H} > 0$$

As human effort cost increases, the numerator rises while the denominator shrinks, both pushing R^* upward. This means firms require even higher retention stickiness to justify hiring a costlier human auditor.
Interpretation: In markets with rising labor costs or constrained staffing, AI becomes more attractive unless retention is extremely high.

(2) Cost of AI Effort C_{AI}

$$\frac{\partial R^*}{\partial C_{AI}} < 0$$

If AI becomes cheaper (as technology improves), C_{AI} decreases, shrinking the numerator and lowering R^*
Interpretation: As AI efficiency increases, the retention bar for human auditors gets even higher, making policy interventions more urgent.

(3) Discount Factor β

$$\frac{\partial R^*}{\partial C_{\beta}} < 0$$

A higher discount factor increases the weight on future periods, enlarging the denominator.
Interpretation: In industries where firms are more future-oriented (high β), the appeal of long-term human retention increases, reducing the required threshold.

(4) Audit Fee F

$$\frac{\partial R^*}{\partial C_{\mathrm{F}}} < 0$$

A higher fee increases the denominator, making retention more profitable.
Interpretation: In high-fee engagements, retention becomes easier to justify economically, even if human effort is more costly.

Table 1. Sensitivity of Critical Retention R^* to Key Inputs

Variable	Sign	Interpretation
C_H	+	Human effort is more expensive → Higher retention needed to justify hiring
C_{AI}	-	AI becomes cheaper → Human auditors need more stickiness to compete
β	-	Firms value future more → Retention becomes more attractive
F	-	Higher audit fees → More room to accommodate human inefficiency

3.2.7. Audit Quality Loss from Retention-Based Selection

Even when the human auditor meets the minimum quality threshold, they do so less efficiently. If a firm hires a human solely because $R > R^*$, society absorbs a quality loss:

$$\Delta.Q = Q_{AI}^* - Q_H^* = Q_{min} - q \,.\, \ln(1+ e_H^*)$$

This gap is not trivial. In practice, even small declines in audit quality may compound into significant market-wide impacts—such as diminished trust, higher capital costs, or increased audit failures. While a degree of retention is economically rational, our model reveals that retention is not free — and should not be unbounded.

This mathematical structure gives regulators a concrete tool to guide retention policy. If retention exceeds R^*, firms may consistently select underperforming auditors. The cost is measurable in audit quality. Policymakers should use this insight to establish retention ceilings, mandate rotation, or require disclosures when retention overrides performance.

This framework does not dismiss human auditors — it explains under what conditions they can be retained without harming audit quality. By understanding the slope and drivers of R^*, we empower firms and regulators to protect both economic efficiency and public trust in the audit system.

IV. POLICY IMPLICATIONS

The behavioral model presented in this paper reveals a structural misalignment between audit performance and audit selection. While AI auditors deliver higher audit quality under fixed fee conditions due to their risk neutrality and effort efficiency, human auditors persist in the market through relationship-based retention. In multi-period engagements, firms may rationally choose to hire a lower-quality human auditor if the expected value of retention outweighs the short-run cost of inefficiency. This creates a measurable trade-off between economic continuity and audit quality integrity, with clear implications for policy design.

Our model introduces a critical retention threshold R^* that determines when human auditors become viable in the presence of AI. This expression depends on human and AI cost structures, the audit fee, and the discount factor. Importantly, we show that even when retention allows a human auditor to be selected, audit quality suffers by a calculable amount. $\Delta.Q = Q_{AI}^* - Q_H^*$. This loss in quality is not theoretical—it is measurable, persistent, and socially relevant. In light of these findings, several policy responses are warranted.

4.1 Set Retention Caps to Protect Audit Quality

The model demonstrates that retention, while economically rational from the firm's perspective, can lead to socially suboptimal audit outcomes. If retention is unbounded, firms may consistently favor relational loyalty over audit performance. Regulators should therefore consider:

- *Retention ceilings*: Establish maximum tenure for audit engagements (e.g., 5–10 years).
- *Rotation mandates*: Require audit firm or partner rotation to limit long-term relationships.
- *Phase-out thresholds*: Implement decreasing retention allowances as AI capabilities improve and cost advantages widen.

These mechanisms limit the over-selection of human auditors when objective quality is lower and prevent structural erosion of audit reliability.

4.2 Increase Transparency in Retention-Driven Engagements

When retention becomes a decisive factor in auditor selection, stakeholders should be informed. We recommend:

- *Disclosure of engagement duration*: Clients and investors deserve visibility into how long an auditor has served a given firm.
- *Retention rationale reporting*: Firms should disclose whether selection was based on technical quality, cost, or relationship history.
- *Retention-risk tagging*: Regulatory bodies can flag audits where retention exceeds a benchmark and quality falls below sector norms.

These steps increase accountability and give the market information to judge auditor independence and quality alignment.

4.3 Encourage Hybrid Audit Structures with Clear Accountability

Rather than forcing a binary choice between human and AI auditors, the market should evolve toward hybrid structures in which:

- AI systems perform the analytical and testing workload, ensuring quality, completeness, and consistency.
- Human auditors retain interpretive judgment and signatory responsibility**,** maintaining legal accountability and relationship management.

This model ensures that audit quality improves without creating a gap in responsibility. It also gives human auditors a path forward—not by competing with machines on processing, but by complementing them through human judgment, communication, and accountability.

4.4 Align Regulatory Strategy with R^* and ΔQ

Our model gives policymakers quantifiable targets:

- The retention threshold R^* defines how sticky human-auditor relationships must be to justify their selection.

- The audit quality loss ΔQ defines how much society pays when the wrong agent is chosen.

These values can be used to:

- Simulate scenarios under varying cost and effort assumptions,
- Set data-informed caps or disclosure requirements,
- And forecast the consequences of unchecked retention policies.

Audit regulation has long relied on principles of independence and objectivity. Our model offers a structural complement to those principles: an analytical framework to measure when relationship-based retention begins to undermine audit performance—and to design limits that restore balance.

5. CONCLUSION

This paper develops a behavioral model comparing the audit quality produced by human and AI auditors in a competitive market. Unlike prior work that treats AI as superior by assumption, we provide a structured explanation grounded in auditor effort, cost, and risk preferences. Human auditors are modeled as risk-averse and effort-averse, while AI systems are risk-neutral and effort-neutral. Holding audit fees constant, we show that AI achieves higher audit quality in a single-period setting, not because of intelligence or innovation, but because it operates without the internal frictions that constrain human effort (Antle & Nalebuff, 1991; DeAngelo, 1981).

However, in a multi-period setting, the economics shift. Human auditors gain a retention advantage—clients are more likely to continue engagements due to familiarity, trust, or inertia (Ashton & Ashton, 1988; Chang & Hwang, 2003). We derive the critical retention threshold R*, above which human auditors become more profitable for firms despite producing lower audit quality. We then quantify the quality loss ΔQ associated with this trade-off and demonstrate how even modest levels of stickiness can distort auditor selection.

Our model contributes to the growing literature on AI adoption in assurance (Appelbaum, Kogan, & Vasarhelyi, 2017) by highlighting not just performance differences, but the conditions under which those differences matter. It provides formal structure to debates on auditor rotation, hybrid audit systems, and accountability in algorithm-assisted assurance (Alles, 2022).

From a policy standpoint, our findings support the case for retention ceilings, enhanced disclosure rules, and accountability frameworks that align selection incentives with audit quality. Retention is not inherently bad—but when it overrides performance, it introduces measurable social costs. The model offers a toolkit for regulators to evaluate the trade-offs, set tolerable thresholds, and design audit systems that balance innovation, continuity, and public trust.

REFERENCES

Alles, M. G. (2022). Audit in the age of artificial intelligence: Navigating automation and accountability. *Accounting Horizons, 36*(2), 35–48.

Annelin, A., & Svanström, T. (2022). The triggers and consequences of audit team stress: Qualitative evidence from engagement teams. *International Journal of Auditing*, *26*(2), 113-133.

Antle, R., & Nalebuff, B. (1991). Conservatism and auditor-client negotiations. *Journal of Accounting Research, 29*, 31–54.

Appelbaum, D., Kogan, A., & Vasarhelyi, M. A. (2017). Big data and analytics in the modern audit engagement: Research needs. *Auditing: A Journal of Practice & Theory, 36*(4), 1–27.

Ashton, R. H., & Ashton, A. H. (1988). Audit delay and the timeliness of earnings announcements. *Journal of Accounting and Economics, 10*(1), 65–83.

Bell, T. B., Landsman, W. R., & Shackelford, D. A. (2001). Auditors' perceived business risk and audit fees: Analysis and evidence. *Journal of Accounting Research, 39*(1), 35–43.

Brown-Liburd, H., Issa, H., & Lombardi, D. (2015). Behavioral implications of Big Data's impact on audit judgment and decision making and future research directions. *Accounting Horizons, 29*(2), 451–468.

Chang, C. J. & Hwang, N. (2003). The impact of retention incentives and client business risks on auditors' decisions involving aggressive reporting practices. *Auditing: A Journal of Practice & Theory, 22*(2), 207–218.

DeAngelo, L. E. (1981). Auditor independence, 'low balling,' and disclosure regulation. *Journal of Accounting and Economics, 3*(2), 113–127.

DeFond, M., & Zhang, J. (2014). A review of archival auditing research. *Journal of Accounting and Economics, 58*(2–3), 275–326.

Fedyk, A., Hodson, J., Khimich, N., & Fedyk, T. (2022). Is artificial intelligence improving the audit process? *Review of Accounting Studies, 27*(3), 938-985.

Francis, J. (2011). A framework for understanding and researching audit quality. *Auditing: A Journal of Practice & Theory, 30*(2), 125–152.

Francis, J., & Yu, M. D. (2009). Big 4 office size and audit quality. *The Accounting Review, 84*(5), 1521–1552.

Ghosh, A., & Siriviriyakul, S. (2018). Quasi rents to audit firms from longer tenure. *Accounting Horizons, 32*(2), 81–102.

Hay, D., Knechel, W. R., & Wong, N. (2006). Audit fees: A meta-analysis of the effect of supply and demand attributes. *Contemporary Accounting Research, 23*(1), 141–191.

Knechel, W. R., & Willekens, M. (2006). The role of risk management and governance in determining audit demand. *Journal of Business Finance & Accounting, 33*(9–10), 1344–1367.

Law, K. K., & Shen, M. (2025). How does artificial intelligence shape audit firms? *Management Science*, *71*(5), 3641-3666.

Lehner, O. M., Ittonen, K., Silvola, H., Ström, E., & Wührleitner, A. (2022). Artificial intelligence based decision-making in accounting and auditing: ethical challenges and normative thinking. *Accounting, Auditing & Accountability Journal*, *35*(9), 109-135.

PwC. (2017). Sizing the prize what's the real value of AI for your business and how can you capitalise? Retrieved May 2 2025 from https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf.

Rennie, M. D., Kopp, L. S., & Lemon, W. M. (2010). Exploring trust and the auditor–client relationship: Factors influencing the auditor's trust of a client representative. *Auditing: A Journal of Practice & Theory, 29*(1), 279–293.

Simunic, D. A. (1980). The pricing of audit services: Theory and evidence. *Journal of Accounting Research, 18*(1), 161–190.

Singer, Z., & Zhang, J. (2018). Auditor tenure and the timeliness of misstatement discovery. *The Accounting Review, 93*(2), 315–338.

Yang, S., Lee, W., Lim, Y., & Yi, C. (2021). Audit firm operating leverage and pricing strategy: Evidence from lowballing in audit industry. *Journal of Contemporary Accounting & Economics, 17*(2), 100254.