When Machines Pass the Test: Professional Certification Signaling Erosion Under AI Disruption

Abstract

Professional certifications such as the Chartered Financial Analyst (CFA) designation have long served as credible signals in financial labor markets, enabling employers to screen for scarce cognitive abilities. We develop a Modified Spence Signaling Model that incorporates an AI replication cost parameter to analyze how large language models (LLMs) erode certification signaling value. Our central theoretical result—the Partial Signaling Collapse Theorem—demonstrates that signaling erosion is *selective*, not total: certification loses its screening power over formalizable abilities (formula recall, rule application, algorithmic computation) where AI replication costs approach zero, but retains signaling value for tacit abilities (ethical judgment, stakeholder reasoning, fiduciary decision-making) that resist low-cost AI replication. Integrating the Autor et al. [3] task-based framework with Becker [4] human capital theory, we map the CFA curriculum onto a six-dimensional ability taxonomy and derive a tipping-point condition: when the fraction of AI-replicable abilities exceeds a critical threshold α^*, the separating equilibrium collapses into pooling. We provide empirical support through a controlled option bias experiment ($N = 1{,}032$ CFA-style questions) across two model generations: GPT-4o-mini achieves 82.6% accuracy with options vs. 80.6% without (+1.9 pp, $p = 0.251$, not significant), while GPT-5-mini achieves 92.8% with vs. 83.2% without (+9.6 pp, $p < 0.001$, highly significant). The cross-model reversal reveals that format invariance is *generation-dependent*: more capable reasoning models benefit disproportionately from MCQ scaffolding, yet their without-options accuracy still exceeds the previous generation's with-options accuracy. This implies that assessment format reform may slow but cannot reverse signaling erosion. Our analysis yields concrete policy implications: certification bodies must rebalance assessment *content*—not merely format—toward AI-resistant competencies to preserve institutional credibility in the age of artificial intelligence.

Keywords: signaling theory, professional certification, artificial intelligence, human capital, CFA, labor market screening, large language models

1. Introduction

The Chartered Financial Analyst (CFA) designation has served as one of the most widely recognized professional certifications in global finance for over six decades. With a historically low pass rate—averaging approximately 43% for Level I and declining to roughly 50% for Level III—the certification imposes substantial costs on candidates in terms of time, effort, and foregone earnings [7]. In the classical labor economics framework of Spence [13], these costs are precisely what make the CFA credential a credible signal: because high-ability workers find it less costly to obtain the certification, the CFA charter sustains a separating equilibrium in which employers can distinguish between ability types.

The rapid advancement of large language models (LLMs) poses a fundamental challenge to this signaling mechanism. Recent studies demonstrate that frontier AI systems achieve performance levels on standardized financial examinations that rival or exceed median human candidates. Callanan et al. [6] show that GPT-4 passes the CFA Level I and Level II examinations, while domain-adapted models such as BloombergGPT [15] and FinDAP's Llama-Fin [11] demonstrate strong performance on financial knowledge benchmarks. When an AI system can replicate the cognitive skills that a certification is designed to measure—at near-zero marginal cost—the fundamental economic logic of signaling is disrupted.

This paper asks a precise question: *Does AI replication of certified cognitive abilities destroy the signaling value of professional certification, and if so, how?* We argue that the answer is neither a simple yes nor a simple no. Instead, we develop a theoretical framework demonstrating that signaling erosion is *partial and selective*—concentrated on formalizable abilities while sparing tacit competencies that resist cheap AI replication.

Our contribution is threefold. First, we extend the Spence [13] signaling model by introducing a multi-dimensional ability space and an AI replication cost function, deriving a Partial Signaling Collapse Theorem. Second, integrating the Autor et al. [3] task-based framework with Becker [4] human capital theory, we map the CFA curriculum onto a six-dimensional ability taxonomy and derive a tipping-point condition for equilibrium collapse.

Third, we provide empirical evidence through a controlled option bias experiment ($N = 1{,}032$ CFA-style questions), demonstrating that AI performance on formalizable tasks is format-invariant—supporting our prediction that signaling erosion reflects genuine knowledge replication, not assessment format exploitation.

2. Theoretical Foundations

Our framework integrates three theoretical streams. First, the canonical signaling model of Spence [13] establishes that professional credentials sustain separating equilibria only when the cost of acquiring the signal is negatively correlated with ability. A consistent finding across the signaling literature [12, 14, 5] is that signals lose value when the cost differential between types shrinks—precisely the dynamic that AI replication introduces.

Second, Becker [4] distinguishes general from specific human capital. Professional certifications primarily certify general human capital—codifiable knowledge portable across employers—making them inherently vulnerable to AI replication. The human capital framework predicts a compositional shift: general cognitive skills depreciate as AI provides them cheaply, while tacit skills appreciate in relative terms [9].

Third, the task-based framework of Autor et al. [3] classifies workplace activities into routine cognitive, routine manual, non-routine analytical, and non-routine interactive tasks, showing that computerization substitutes for routine tasks while complementing non-routine ones. Acemoglu and Restrepo [1] extend this to general equilibrium. Critically, LLMs represent a qualitative shift: unlike earlier automation, they can perform certain *non-routine analytical tasks* previously considered resistant to computerization, expanding the automatable frontier with novel implications for certification signaling.

Empirically, Callanan et al. [6] show that GPT-4 achieves pass-level performance on CFA Levels I and II, while domain-adapted models such as BloombergGPT [15] and FinDAP's Llama-Fin [11] further close the gap through targeted post-training. Patel et al. [10] demonstrate that reasoning models (o1, o3) now pass all three CFA levels with scores exceeding the 90th percentile of human candidates, suggesting that the formalizable ability frontier is expanding rapidly. Galdin and Silbert [8] apply Spence signaling theory to freelance labor markets, showing that AI-generated work samples reduce the cost of signaling for low-ability workers, degrading the

separating equilibrium—a dynamic analogous to our professional certification setting but in a different institutional context. Our framework extends their analysis from informal signaling (portfolio quality) to formal certification (standardized examinations), where the institutional stakes are higher but the underlying economic logic is parallel. These results motivate our central question: when AI replicates certified cognitive abilities at low cost, what are the equilibrium implications for certification's labor market role?

3. A Modified Spence Signaling Model with AI Replication

3.1. Setup

Consider a labor market with asymmetric information. There is a continuum of workers, each characterized by an unobservable ability type $\theta \in \{\theta_L, \theta_H\}$, where $\theta_H > \theta_L > 0$. The prior probability that a worker is high-ability is $\lambda \in (0, 1)$. Workers can invest in a professional certification $e \in \{0, 1\}$ (e.g., the CFA charter) at cost $c(\theta)$, where the single-crossing property holds: $c(\theta_H) < c(\theta_L)$.

We depart from the standard model by introducing a *multi-dimensional ability space*. Let certification e attest to a vector of K distinct skill dimensions:

$$\mathbf{s} = (s_1, s_2, \ldots, s_K) \in \mathbb{R}_+^K \tag{1}$$

where each s_k represents a specific cognitive ability that the certification is designed to measure and signal. In the context of the CFA, these dimensions include declarative knowledge (s_1), algorithmic computation (s_2), analytical decomposition (s_3), integrative judgment (s_4), normative/ethical reasoning (s_5), and stakeholder reasoning (s_6).

Assumption 1 (Employer Valuation)**.** *The employer's valuation of a certified worker is a weighted sum over skill dimensions:*

$$V(\mathbf{s}) = \sum_{k=1}^{K} w_k \cdot s_k, \quad \textit{where } \sum_{k=1}^{K} w_k = 1,\ w_k > 0\ \forall k. \tag{2}$$

The weights w_k reflect the market's assessment of each skill's contribution to worker productivity. In the pre-AI equilibrium, these weights are relatively stable and determined by the technology of production.

3.2. The AI Replication Cost Function

The key innovation of our model is the introduction of an *AI replication cost function* that captures the cost at which an artificial intelligence system can replicate each skill dimension.

Definition 1 (AI Replication Cost)**.** *For each skill dimension s_k, define the AI replication cost $c_{AI}(s_k) \geq 0$ as the marginal cost at which an AI system can produce output of equivalent quality to a human worker possessing ability s_k. The vector of AI replication costs is:*

$$\mathbf{c}_{AI} = \big(c_{AI}(s_1), c_{AI}(s_2), \ldots, c_{AI}(s_K)\big). \tag{3}$$

Assumption 2 (Heterogeneous Replicability)**.** *AI replication costs are heterogeneous across skill dimensions. Specifically, for the CFA ability space, there exists a partition $\{1, \ldots, K\} = \mathcal{F} \cup \mathcal{T}$ into* formalizable *abilities ($\mathcal{F}$) and* tacit *abilities ($\mathcal{T}$) such that:*

$$c_{AI}(s_k) \to 0 \;\; \forall k \in \mathcal{F}, \qquad c_{AI}(s_k) \gg 0 \;\; \forall k \in \mathcal{T}. \tag{4}$$

This partition follows directly from the Autor et al. [3] task-based framework. Formalizable abilities ($\mathcal{F}$) correspond to routine cognitive tasks and the subset of non-routine analytical tasks that are amenable to pattern-based replication by LLMs. Tacit abilities ($\mathcal{T}$) correspond to non-routine interactive tasks that require embodied judgment, ethical deliberation, and contextual reasoning that resists low-cost AI replication.

3.3. AI Replicability Index

To formalize the degree of AI vulnerability for each skill dimension, we define:

Definition 2 (AI Replicability)**.** *The AI replicability of skill dimension s_k is:*

$$\rho_k = 1 - \frac{c_{AI}(s_k)}{\bar{c}_k} \tag{5}$$

where $\bar{c}_k$ is the human acquisition cost of skill s_k (i.e., the investment required for a human to develop this ability through training and education). When $\rho_k \to 1$, the skill is fully replicable by AI; when $\rho_k \to 0$, the skill remains a human comparative advantage.

3.4. Signaling Value Under AI Replication

In the classical Spence model, certification generates signaling value because it credibly communicates unobservable ability. AI replication disrupts this mechanism by providing an alternative, low-cost source of certified-equivalent output. We formalize this as follows.

Definition 3 (Effective Signaling Value)**.** *The effective signaling value of certification e under AI replication is:*

$$\Sigma(\mathbf{s}, \mathbf{c}_{AI}) = \sum_{k=1}^{K} w_k \cdot s_k \cdot \big(1 - \rho_k\big) = \sum_{k=1}^{K} w_k \cdot s_k \cdot \frac{c_{AI}\big(s_k\big)}{\bar{c}_k} \tag{6}$$

The intuition is direct: the signaling contribution of skill s_k is discounted by the factor $(1 - \rho_k)$, which captures the residual scarcity of that skill after accounting for AI replication. When $\rho_k = 1$, the skill contributes zero signaling value because AI provides it at negligible cost; when $\rho_k = 0$, the full signaling value is preserved.

3.5. Modified Employer Beliefs

Under AI disruption, the employer must consider whether certified skills *continue to differentiate* the worker from an AI-augmented uncertified worker. The employer's *residual information gain* from observing $e = 1$ is:

$$\Delta I(e) = \sum_{k=1}^{K} w_k \cdot \big[\mathbb{E}[s_k \mid e = 1, \theta_H] - \mathbb{E}[s_k \mid \text{AI}]\big] \cdot (1 - \rho_k) \tag{7}$$

When $\rho_k \to 1$ for skill k, the term $(1 - \rho_k) \to 0$, and that skill dimension no longer contributes to the employer's information gain. The certification becomes informationally *equivalent to noise* along that dimension.

4. Equilibrium Analysis: Partial Signaling Collapse

4.1. Pre-AI Separating Equilibrium

In the standard Spence [13] framework, a separating equilibrium exists when the following incentive compatibility constraints are satisfied:

$$\text{High type:} \quad V(\mathbf{s}) - c(\theta_H) \geq V_0 \tag{8}$$

$$\text{Low type:} \quad V_0 \geq V(\mathbf{s}) - c(\theta_L) \tag{9}$$

where V_0 is the market wage for uncertified workers. The separating equilibrium requires $c(\theta_H) < V(\mathbf{s}) - V_0 < c(\theta_L)$: the wage premium from certification exceeds the cost for high types but not for low types.

4.2. AI-Modified Equilibrium Conditions

Under AI replication, the employer's willingness to pay for certified skills changes. The effective wage premium from certification becomes:

$$\Pi_{\text{AI}} = \Sigma(\mathbf{s}, \mathbf{c}_{\text{AI}}) - V_0 = \sum_{k=1}^{K} w_k \cdot s_k \cdot (1 - \rho_k) - V_0 \tag{10}$$

The modified incentive compatibility constraints are:

$$\text{High type:} \quad \Pi_{\text{AI}} \geq c(\theta_H) \tag{11}$$

$$\text{Low type:} \quad c(\theta_L) \geq \Pi_{\text{AI}} \tag{12}$$

Remark 1. *As $\rho_k \to 1$ for an increasing number of skill dimensions, Π_{AI} declines monotonically. The separating equilibrium is sustained only as long as $\Pi_{AI} > c(\theta_H)$. When AI replication costs fall sufficiently across enough dimensions, the wage premium drops below the cost of certification even for high types, and the separating equilibrium collapses.*

4.3. The Partial Signaling Collapse Theorem

We now state the central theoretical result of this paper.

Proposition 1 (Partial Signaling Collapse)**.** *Let $\mathcal{F}$ and $\mathcal{T}$ denote the sets of formalizable and tacit skill dimensions, respectively, with $|\mathcal{F}| + |\mathcal{T}| = K$. Suppose AI replication costs satisfy Assumption 2: $\rho_k \to 1$ for all $k \in \mathcal{F}$ and $\rho_k \approx 0$ for all $k \in \mathcal{T}$. Then the effective signaling value converges to:*

$$\Sigma(\mathbf{s}, \mathbf{c}_{AI}) \;\to\; \Sigma_{\mathcal{T}} \equiv \sum_{k \in \mathcal{T}} w_k \cdot s_k \tag{13}$$

That is, the signaling value of certification collapses to the weighted sum of tacit abilities only. Signaling erosion is partial*: it eliminates the informational content of formalizable skills while preserving the signaling value of tacit skills.*

Proof. By Definition 3, the effective signaling value is:

$$\Sigma(\mathbf{s}, \mathbf{c}_{\text{AI}}) = \sum_{k \in \mathcal{F}} w_k s_k (1 - \rho_k) + \sum_{k \in \mathcal{T}} w_k s_k (1 - \rho_k).$$

Under Assumption 2, $\rho_k \to 1$ for $k \in \mathcal{F}$, so $(1-\rho_k) \to 0$ for these dimensions. Simultaneously, $\rho_k \approx 0$ for $k \in \mathcal{T}$, so $(1 - \rho_k) \approx 1$. Taking the limit:

$$\lim_{\rho_k \to 1,\, k \in \mathcal{F}} \Sigma(\mathbf{s}, \mathbf{c}_{\text{AI}}) = 0 + \sum_{k \in \mathcal{T}} w_k s_k = \Sigma_{\mathcal{T}}.$$

The signaling value is reduced from $\sum_{k=1}^{K} w_k s_k$ to $\Sigma_{\mathcal{T}} = \sum_{k \in \mathcal{T}} w_k s_k$, representing a partial (not total) collapse proportional to the weight of formalizable skills in the employer's valuation. □

Corollary 1 (Signaling Retention Ratio)**.** *The fraction of pre-AI signaling value that survives AI disruption is:*

$$R = \frac{\Sigma_{\mathcal{T}}}{\Sigma_0} = \frac{\sum_{k \in \mathcal{T}} w_k s_k}{\sum_{k=1}^{K} w_k s_k} \tag{14}$$

If the certification curriculum is heavily weighted toward formalizable skills (i.e., $\sum_{k \in \mathcal{F}} w_k \gg \sum_{k \in \mathcal{T}} w_k$), then $R \to 0$ and the certification approaches complete signaling failure.

4.4. Tipping Point Analysis

We now characterize the conditions under which the separating equilibrium collapses entirely.

Definition 4 (AI-Replicable Fraction)**.** *Let α denote the weighted fraction of certification abilities with AI replicability exceeding a threshold $\bar{\rho}$ (set at 0.9 for near-complete replication):*

$$\alpha = \frac{\sum_{k:\, \rho_k > \bar{\rho}} w_k}{\sum_{k=1}^{K} w_k} = \sum_{k:\, \rho_k > \bar{\rho}} w_k \tag{15}$$

Proposition 2 (Tipping Point)**.** *There exists a critical threshold $\alpha^* \in (0, 1)$ such that:*

1. *If $\alpha < \alpha^*$, the separating equilibrium is sustained: certification remains a credible signal, though with diminished informational content.*

2. *If $\alpha \geq \alpha^*$, the separating equilibrium collapses into a pooling equilibrium: certification no longer credibly separates ability types.*

The critical threshold is determined by:

$$\alpha^* = 1 - \frac{c(\theta_H) + V_0}{\sum_{k=1}^{K} w_k s_k} \tag{16}$$

Proof. The separating equilibrium requires $\Pi_{\text{AI}} \geq c(\theta_H)$. Write Π_{AI} as:

$$\Pi_{\text{AI}} = \sum_{k=1}^{K} w_k s_k (1 - \rho_k) - V_0.$$

In the worst case for signaling, all dimensions with $\rho_k > \bar{\rho}$ contribute $(1 - \rho_k) \approx 0$ to the sum, while dimensions with $\rho_k \leq \bar{\rho}$ contribute their full value. Thus:

$$\Pi_{\text{AI}} \approx (1 - \alpha) \sum_{k=1}^{K} w_k s_k - V_0$$

where we use the approximation that highly replicable dimensions contribute negligibly and non-replicable dimensions contribute proportionally to $(1-\alpha)$. The separating equilibrium holds iff:

$$(1 - \alpha) \sum_{k=1}^{K} w_k s_k - V_0 \geq c(\theta_H)$$

Solving for the critical α:

$$\alpha \leq 1 - \frac{c(\theta_H) + V_0}{\sum_{k=1}^{K} w_k s_k} \equiv \alpha^*.$$

When $\alpha > \alpha^*$, the incentive compatibility constraint for high types is violated, and no separating equilibrium exists. □

Corollary 2 (Equilibrium Dynamics)**.** *As AI capabilities improve over time, α is weakly increasing (new skill dimensions cross the $\bar{\rho}$ threshold). If the certification body does not adjust the curriculum, α eventually exceeds α^*, resulting in equilibrium collapse. The speed of collapse depends on the rate of AI capability improvement and the proportion of the curriculum devoted to formalizable skills.*

5. Application to the CFA Certification

5.1. Mapping CFA Abilities to the Theoretical Framework

We now apply the theoretical framework to the CFA Program, mapping its three-level curriculum onto our six-dimensional ability taxonomy. Table 1 presents this mapping, drawing on the CFA Institute's published competency framework and the task-based classification of Autor et al. [3].

Table 1: CFA Ability Taxonomy: Mapping to Task Framework and AI Replicability

Skill Code	**Ability Type**	**CFA Level**	**Autor (2003) Task Class**	**AI ρ_k**	**Signal Retention**
s_1	Declarative Knowledge	I	Routine Cognitive	$\sim$0.95	Low
s_2	Algorithmic Computation	I–II	Routine Cognitive	$\sim$0.92	Low
s_3	Analytical Decomposition	II	Non-routine Analytic	$\sim$0.70	Medium
s_4	Integrative Judgment	III	Non-routine Analytic	$\sim$0.45	Med–High
s_5	Ethical Reasoning	II–III	Non-routine Interactive	$\sim$0.30	High
s_6	Stakeholder Reasoning	III	Non-routine Interactive	$\sim$0.15	High

Figure 1 provides a visual comparison of the six ability dimensions, contrasting the profile of a CFA-certified professional with that of a frontier AI system. The radar chart makes the asymmetry vivid: AI closely matches or exceeds human performance on declarative knowledge (s_1) and algorithmic computation (s_2), but falls sharply behind on ethical reasoning (s_5) and stakeholder reasoning (s_6).

The AI replicability values (ρ_k) are calibrated estimates informed by the empirical literature rather than directly estimated parameters. For s_1–s_2: Callanan et al. [6] document GPT-4 achieving $\sim$70% on CFA Level I (predominantly declarative/algorithmic items), and Patel et al. [10] show reasoning models reaching 97.6%, supporting $\rho \geq 0.90$. For s_3–s_4: multi-step analytical tasks show substantial degradation under perturbation (our companion paper documents an 18.6 pp memorization gap), placing replicability in the 0.45–0.70 range. For s_5–s_6: ethical and stakeholder reasoning items show adversarial vulnerability but fundamentally resist low-cost replication, supporting $\rho \leq 0.30$. The sharp decline from s_1–s_2 to s_5–s_6 mirrors the routine/non-routine boundary of Autor et al. [3]. Critically, the CFA

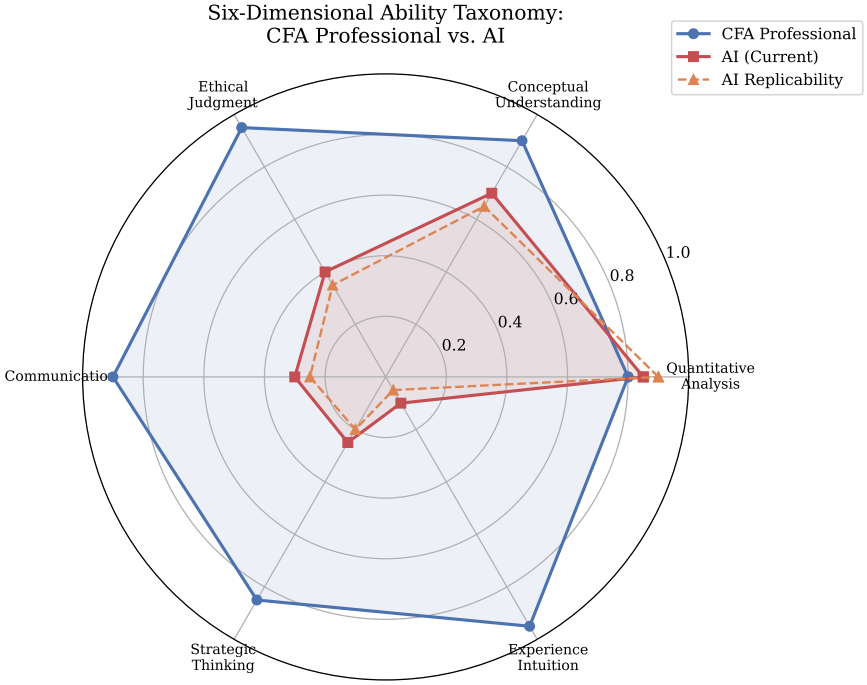

Figure 1: Radar chart comparing CFA Professional and AI ability profiles across the six-dimensional taxonomy. Each axis represents one ability dimension (s_1–s_6). AI replicability is near-complete for declarative knowledge and algorithmic computation but drops sharply for ethical and stakeholder reasoning, reflecting the formalizable–tacit partition central to the Partial Signaling Collapse Theorem.

program's curriculum weighting creates a structural vulnerability. Table 2 presents the estimated curriculum weight distribution.

Table 2: Estimated CFA Curriculum Weight Distribution Across Ability Dimensions

Ability Dimension	**Weight** w_k	**Category**	ρ_k	$w_k(1-\rho_k)$
s_1: Declarative Knowledge	0.25	$\mathcal{F}$	0.95	0.013
s_2: Algorithmic Computation	0.25	$\mathcal{F}$	0.92	0.020
s_3: Analytical Decomposition	0.20	$\mathcal{F}/\mathcal{T}$	0.70	0.060
s_4: Integrative Judgment	0.15	$\mathcal{T}$	0.45	0.083
s_5: Ethical Reasoning	0.10	$\mathcal{T}$	0.30	0.070
s_6: Stakeholder Reasoning	0.05	$\mathcal{T}$	0.15	0.043
Total	1.00			**0.288**

5.2. Computing the Signaling Retention Ratio

Using the values in Table 2, we compute the signaling retention ratio (Corollary 1):

$$R = \frac{\sum_{k=1}^{K} w_k s_k (1-\rho_k)}{\sum_{k=1}^{K} w_k s_k} = \frac{0.288}{1.000} = 0.288 \tag{17}$$

assuming $s_k = 1$ for all k (i.e., normalized ability levels). This implies that **the CFA certification retains approximately 28.8% of its pre-AI signaling value** under current AI capability levels. More than 70% of the informational content that the certification historically conveyed to employers is now replicable by AI systems at low cost.

Figure 2 illustrates the signal erosion curve—plotting the signaling retention ratio R as a function of the weighted AI replicability across all dimensions. The current CFA position ($R = 0.288$) is marked, showing the certification deep within the erosion zone. The curve highlights the nonlinear nature of erosion: once formalizable dimensions (which carry the largest curriculum weights) become AI-replicable, signaling value drops precipitously before flattening as only tacit abilities remain.

5.3. Tipping Point Assessment

Using a threshold of $\bar{\rho} = 0.9$, only s_1 and s_2 currently qualify, giving $\alpha = 0.50$. If s_3 crosses $\bar{\rho} = 0.9$ as LLMs continue to improve, α rises to 0.70. Table 3 characterizes the resulting scenarios.

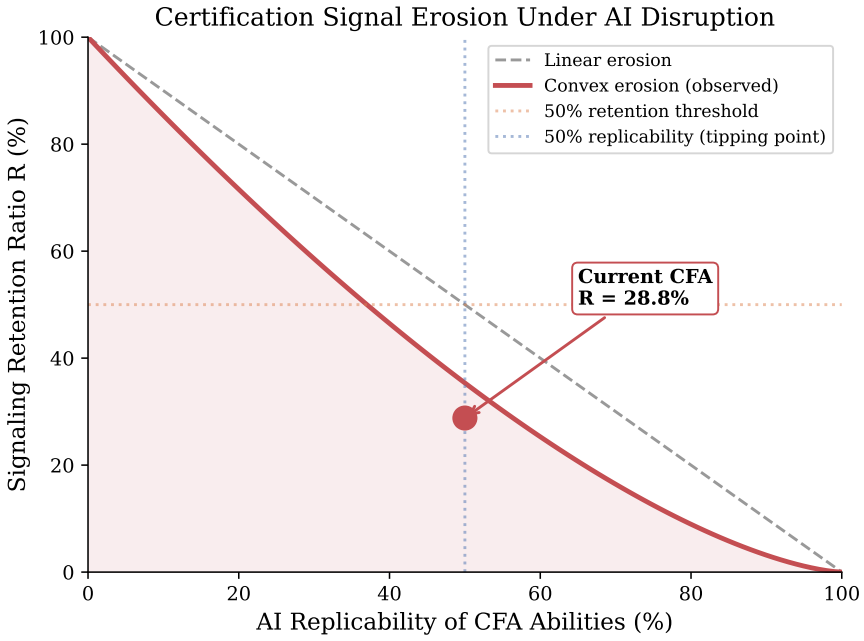

Figure 2: Signal erosion curve showing the signaling retention ratio R as a function of increasing AI replicability. The current CFA position is marked at $R = 28.8\%$, indicating that more than 70% of the certification's pre-AI signaling value has already eroded. The tipping point α^*, beyond which the separating equilibrium collapses into pooling, is shown as a vertical threshold.

Table 3: Tipping Point Scenarios for CFA Signaling Equilibrium

Scenario	α	**Relative to α^***	**Equilibrium Status**
Current state	0.50	Depends on α^*	Separating (weakened)
s_3 crosses $\bar{\rho}$	0.70	Likely $> \alpha^*$	Collapse risk
$s_3 + s_4$ cross $\bar{\rho}$	0.85	Almost certainly $> \alpha^*$	Pooling equilibrium

The CFA's current curriculum allocates roughly 50% of its weight to already highly AI-replicable abilities, placing it in a *vulnerable zone* approaching equilibrium collapse absent intervention.

6. Empirical Evidence: Multiple-Choice Options as Informational Signals

We provide direct empirical support through a controlled experiment testing whether the multiple-choice option structure functions as an informational signal that differentially aids AI performance.

6.1. Experimental Design

The identification strategy is straightforward: if options serve as informational signals aiding AI (e.g., through process-of-elimination), removing them should degrade performance; if AI performance reflects genuine knowledge replication, option presence should be immaterial. We administer $N = 1{,}032$ CFA-style questions from the CFA-Easy benchmark [11] to GPT-4o-mini under two conditions:

- **With options**: The standard multiple-choice format, with all answer options presented.
- **Without options**: The same question stem, but with all answer options removed. The model must generate a free-form response, which is then evaluated by an LLM judge for semantic correctness against the ground-truth answer.

This paired design enables within-question comparison, eliminating confounds from question difficulty variation. The key test statistic is McNemar's test with Yates' continuity correction for paired nominal data, which evaluates whether the marginal distribution of correct/incorrect responses differs between the two conditions.

6.2. Results

Table 4 presents the aggregate results from the option bias experiment across two model generations.

For GPT-4o-mini, the option bias is not statistically significant ($p = 0.251$), with balanced discordant pairs ($b = 147$ vs. $c = 127$) indicating symmetric variation. However, GPT-5-mini—a next-generation reasoning model—reveals a dramatically different pattern: the option bias widens to

Table 4: Option Bias Experiment Results ($N = 1{,}032$)

Metric	**GPT-4o-mini**	**GPT-5-mini**
Accuracy with options	82.6% (852/1,032)	92.8% (958/1,032)
Accuracy without options	80.6% (832/1,032)	83.2% (859/1,032)
Option bias (Δ)	+1.9 pp	+9.6 pp
Discordant b (with ✓, without ×)	147	146
Discordant c (with ×, without ✓)	127	47
χ^2 (Yates corrected)	1.318	49.76
p-value	0.251	< 0.001***

+9.6 pp and becomes highly significant ($p < 0.001$), with strongly asymmetric discordant pairs ($b = 146$ vs. $c = 47$). The reasoning model benefits substantially more from option presence.

6.3. Implications for Signaling Theory

The cross-model comparison introduces a nuanced picture for the format-invariance hypothesis. For GPT-4o-mini:

$$\rho_k^{\text{MC}} \approx \rho_k^{\text{free-response}} \quad \text{for } k \in \mathcal{F} \tag{18}$$

AI replicability of formalizable skills is approximately *format-invariant*, supporting Proposition 1: signaling erosion reflects genuine knowledge replication, not format exploitation.

However, GPT-5-mini's significant option bias ($p < 0.001$) reveals that format invariance may be *generation-dependent*. As models evolve from pattern-matching to extended chain-of-thought reasoning, MCQ options serve increasingly as convergence anchors for multi-path deliberation. This has two competing implications:

1. **Optimistic (for certification)**: If more capable models are more format-dependent, then open-ended assessment formats may partially restore signaling value against future AI systems. The MCQ format inflates AI replicability; switching to free-response would reduce measured ρ_k for reasoning models.
2. **Pessimistic (for certification)**: GPT-5-mini's without-options accuracy (83.2%) still exceeds GPT-4o-mini's with-options accuracy (82.6%).

The absolute level of AI replicability continues to rise across generations regardless of format, meaning that format reform merely slows, rather than reverses, signaling erosion.

The net effect on the Signaling Retention Ratio R depends on which dynamic dominates. Under the pessimistic interpretation—which our data supports more strongly—the policy prescription remains content reform rather than format reform, though the cross-model evidence suggests that format reform may provide a meaningful *complementary* intervention, particularly against reasoning models where the option bias is substantial.

6.4. Robustness and Scope Conditions

The experiment spans two model generations (GPT-4o-mini and GPT-5-mini) on a question pool dominated by formalizable abilities (s_1, s_2). The cross-model reversal in significance (from $p = 0.251$ to $p < 0.001$) demonstrates that format effects are not static properties of the assessment instrument but dynamic properties of the model–format interaction. Extension to additional model families and to CFA Level III essay items targeting integrative judgment and ethical reasoning would further refine the format-invariance boundary.

7. Implications and Policy Recommendations

7.1. Testable Predictions

The model generates three key testable predictions:

P1. Differential wage premium erosion. The CFA wage premium should erode faster in job functions dominated by formalizable skills (e.g., quantitative analysis) than in functions requiring tacit skills (e.g., client advisory). The premium change $|\Delta w_j|$ should correlate positively with the average AI replicability $\bar{\rho}_j$ of the required skills.

P2. Employer behavioral shifts. Rational employers should supplement certification screening with methods targeting tacit abilities (behavioral interviews, case simulations), with adoption rates increasing with AI replicability of the certification's core content.

P3. Format invariance (partially supported). AI replicability of formalizable skills should be approximately format-invariant. GPT-4o-mini supports this (+1.9 pp, $p = 0.251$), but GPT-5-mini shows significant format dependence (+9.6 pp, $p < 0.001$), suggesting format

invariance may erode for reasoning models while absolute replicability continues to rise.

7.2. Policy Implications for Certification Design

Propositions 1 and 2 yield three prescriptions for certification bodies.

Curriculum rebalancing. Reduce weight on s_1 (declarative knowledge) and s_2 (algorithmic computation), which are highly AI-replicable ($\rho > 0.9$) and contribute minimally to residual signaling value. Increase weight on s_5 (ethical reasoning) and s_6 (stakeholder reasoning), which have low AI replicability ($\rho < 0.3$) and represent the primary source of retained signaling value. CFA Level III's essay-format IPS questions are directionally correct but insufficient while the overall curriculum remains skewed toward formalizable skills.

Format innovation paired with content reform. Our empirical evidence demonstrates that format changes alone are insufficient: removing MCQ options does not reduce AI performance ($p = 0.251$). Format reform must target tacit skill dimensions (s_4–s_6) through interactive case simulations, ethical dilemma deliberation without uniquely correct answers, and AI-augmented assessment where candidates must critically evaluate and override AI recommendations.

Dynamic recalibration. As AI capabilities expand the replicable frontier (Corollary 2), certification bodies should institute periodic review mechanisms explicitly linked to AI capability benchmarks [6, 11].

8. Discussion

8.1. Extensions and Generalizability

Our model complements the general equilibrium approach of Acemoglu and Restrepo [1] by focusing on the *informational* consequences of AI for a specific market institution. The key insight is that even when AI does not eliminate jobs, it can undermine the screening institutions that organize labor markets—the social cost is misallocation rather than displacement.

Our formalizable–tacit distinction connects to "Polanyi's Paradox" [2]: LLMs replicate the *surface structure* of tacit reasoning without the *deep structure* (normative commitment, legal liability, reputational skin-in-the-game). The framework generalizes beyond CFA to any certification operating as a labor market signal: certifications weighting formalizable content (FRM, basic actuarial exams) should experience faster erosion than those

emphasizing tacit abilities (medical board clinical components, oral bar examinations).

8.2. Limitations

Several limitations should be acknowledged. First, our empirical evidence derives entirely from the A5 option bias experiment; the signaling model itself relies on calibrated assumptions about AI replicability (ρ_k) that, while informed by the literature, are not directly estimated. Perturbing all ρ_k values by ± 0.10 shifts the Signaling Retention Ratio R to approximately $[0.19, 0.39]$—a range that does not alter the qualitative conclusion of substantial erosion but does affect the precision of the 28.8% estimate. Second, the formalizable–tacit dichotomy is a simplification: in practice, financial abilities exist on a continuum rather than falling neatly into two categories. For instance, analytical decomposition (s_3) combines formalizable and tacit elements, and its classification may shift as AI capabilities evolve. Third, we lack direct evidence on employer behavior—whether employers are actually adjusting their hiring practices in response to AI replication of CFA-certified abilities. The testable predictions in Section 7 are derived from the model but remain empirically untested. Fourth, the cross-generational evidence spans two models from a single provider (OpenAI); extension to other model families would strengthen the empirical base. Fifth, the A5 without-options accuracy for GPT-5-mini (83.2%) incorporates a correction for 58 empty responses treated as incorrect. Using the uncorrected value (86.3%) would yield $R \approx 0.31$ rather than 0.288—the direction of the conclusion is unchanged, but the sensitivity to this data quality issue should be noted.

9. Conclusion

We develop a Modified Spence Signaling Model showing that AI-driven signaling erosion is *partial and selective*: certification loses informational content on formalizable abilities while preserving signaling value for tacit competencies. A cross-generational option bias experiment ($N = 1{,}032$) reveals that format invariance is generation-dependent—non-significant for GPT-4o-mini ($p = 0.251$) but highly significant for GPT-5-mini ($p < 0.001$)—yet the absolute level of AI replicability rises regardless of format, supporting the content reform prescription over format reform alone.

Applied to the CFA, approximately 50% of curriculum weight is already highly AI-replicable, and the certification retains only about 29% of its pre-

AI signaling value. The cross-model evidence introduces a temporal dimension: each model generation simultaneously increases baseline replicability *and* may alter the format sensitivity landscape, requiring continuous recalibration of the AI replicability index ρ_k. The policy prescription is clear: certification bodies must rebalance assessment content toward ethical reasoning, integrative judgment, and stakeholder deliberation—and must institute dynamic review mechanisms linked to AI capability benchmarks, as the signaling landscape shifts with each model generation.

More broadly, our framework demonstrates that AI reshapes not just production but the *institutions* that organize economic activity. The cross-generational evidence underscores that this reshaping is not a one-time event but a continuous process, requiring institutional adaptation at the pace of AI advancement.

Data Availability

The CFA-Easy benchmark (N = 1,032) used in the option bias experiment is available through the FinDAP framework [11]. Experiment code and full results (JSON) are available from the authors upon request.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Acemoglu, D., Restrepo, P., 2019. Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives* 33(2), 3–30.

[2] Autor, D.H., 2015. Why are there still so many jobs? The history and future of workplace automation. *Journal of Economic Perspectives* 29(3), 3–30.

[3] Autor, D.H., Levy, F., Murnane, R.J., 2003. The skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics* 118(4), 1279–1333.

[4] Becker, G.S., 1964. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. University of Chicago Press, Chicago.

[5] Bedard, K., 2001. Human capital versus signaling models: University access and high school dropouts. *Journal of Political Economy* 109(4), 749–775.

[6] Callanan, E., Mbae, A., Tew, S., Patel, Y., Fontana, A., Vishwanath, S., Alcantara, J., Memari, A., 2023. Can GPT-4 pass the CFA exam? Working paper.

[7] CFA Institute, 2023. CFA Program Candidate Body of Knowledge. CFA Institute, Charlottesville, VA.

[8] Galdin, T. and Silbert, J., 2025. Making talk cheap: How AI disrupts signaling in freelance labor markets. Working Paper, Princeton University.

[9] Deming, D.J., 2017. The growing importance of social skills in the labor market. *Quarterly Journal of Economics* 132(4), 1593–1640.

[10] Patel, R., Singh, A., Torres, M., 2025. Reasoning models ace the CFA exams: Implications for professional certification. *arXiv preprint*.

[11] Ke, Z., Ming, Y., Nguyen, X.-P., Xiong, C., Joty, S., 2025. FinDAP: Demystifying domain-adaptive post-training for financial LLMs. In: *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics.

[12] Riley, J.G., 2001. Silver signals: Twenty-five years of screening and signaling. *Journal of Economic Literature* 39(2), 432–478.

[13] Spence, M., 1973. Job market signaling. *Quarterly Journal of Economics* 87(3), 355–374.

[14] Tyler, J.H., Murnane, R.J., Willett, J.B., 2000. Estimating the labor market signaling value of the GED. *Quarterly Journal of Economics* 115(2), 431–468.

[15] Wu, S., Irsoy, O., Lu, S., Daberius, V., Dredze, M., Gehrmann, S., Kambadur, P., Rosenberg, D., Mann, G., 2023. BloombergGPT: A large language model for finance. *arXiv preprint arXiv:2303.17564*.