

當機器通過考試：AI 衝擊下專業證照訊號侵蝕之研究

Wei-Lun Cheng^a, Daniel Wei-Chung Miao^{a,*}, Guang-Di Chang^a

^a 國立臺灣科技大學財務金融研究所, 臺北, 10607, 臺灣

Abstract

特許金融分析師（CFA）等專業證照長期以來在金融勞動市場中扮演可信賴的訊號角色，使雇主得以篩選稀缺的認知能力。本文建構一個修正式 Spence 訊號模型，引入 AI 複製成本參數，分析大型語言模型（LLMs）如何侵蝕證照訊號價值。我們的核心理論成果——部分訊號崩解定理——證明訊號侵蝕具有選擇性而非全面性：證照在可形式化能力（公式記憶、規則應用、演算法運算）方面喪失篩選功能，因為 AI 複製成本趨近於零；但在隱性能力（倫理判斷、利害關係人推理、受託人決策）方面仍保有訊號價值，因為這些能力抗拒低成本的 AI 複製。我們整合 Autor et al. [3] 任務導向架構與 Becker [4] 人力資本理論，將 CFA 課程對應至六維能力分類體系，並推導臨界點條件：當 AI 可複製能力的比例超過臨界門檻 α^* 時，分離均衡將崩解為混同均衡。我們透過一項控制組選項偏誤實驗（ $N = 1,032$ 道 CFA 風格題目）提供實證支持，橫跨兩個模型世代：GPT-4o-mini 在有選項條件下達到 82.6% 正確率，無選項條件下為 80.6%（+1.9 pp， $p = 0.251$ ，不顯著）；GPT-5-mini 在有選項條件下達到 92.8%，無選項條件下為 83.2%（+9.6 pp， $p < 0.001$ ，高度顯著）。跨模型的顯著性反轉揭示格式不變性具有世代依賴性：推理能力更強的模型從選擇題結構中獲益更多，但其無選項正確率仍高於前一代的有選項正確率。這意味著評量格式改革或可延緩但無法逆轉訊號侵蝕。我們的分析提出具體政策建議：證照機構必須重新調整評量內容——而非僅調整格式——朝向 AI 抗性能力傾斜，以在人工智慧時代維護制度公信力。

Keywords: 訊號理論, 專業證照, 人工智慧, 人力資本, CFA, 勞動市場篩選,

*通訊作者

Email addresses: d11018003@mail.ntust.edu.tw (Wei-Lun Cheng),
miao@mail.ntust.edu.tw (Daniel Wei-Chung Miao), gchang@mail.ntust.edu.tw
(Guang-Di Chang)

1. 緒論

特許金融分析師（CFA）證照六十餘年來一直是全球金融界最受認可的專業資格之一。其歷史通過率極低——Level I 平均約 43%，Level III 更降至約 50%——對考生在時間、精力與機會成本方面構成可觀的負擔 [7]。在 Spence [13] 的古典勞動經濟學架構中，這些成本正是使 CFA 證照成為可信賴訊號的關鍵：因為高能力者取得證照的成本較低，CFA 特許狀得以維繫一個分離均衡，使雇主能夠區分不同能力類型的求職者。

大型語言模型（LLMs）的快速發展對此訊號機制構成根本性挑戰。近期研究顯示，前沿 AI 系統在標準化金融考試中的表現已可與中位數人類考生匹敵甚至超越。Callanan et al. [6] 證明 GPT-4 能通過 CFA Level I 和 Level II 考試，而領域適配模型如 BloombergGPT [15] 與 FinDAP 的 Llama-Fin [11] 則透過針對性後訓練在金融知識基準測試上展現強勁表現。當 AI 系統能以近乎零邊際成本複製證照所欲衡量的認知能力時，訊號的基本經濟邏輯便遭到顛覆。

本文提出一個精確的問題：*AI* 對受認證認知能力的複製是否摧毀了專業證照的訊號價值？若是，其機制為何？我們主張答案既非單純肯定亦非單純否定。相反地，我們發展一個理論架構，證明訊號侵蝕具有部分性與選擇性——集中於可形式化能力，同時保留抗拒廉價 AI 複製的隱性能力。

本文貢獻有三。第一，我們擴展 Spence [13] 訊號模型，引入多維能力空間與 AI 複製成本函數，推導出部分訊號崩解定理。第二，整合 Autor et al. [3] 任務導向架構與 Becker [4] 人力資本理論，將 CFA 課程對應至六維能力分類體系，並推導均衡崩解的臨界點條件。第三，我們透過一項控制組選項偏誤實驗（ $N = 1,032$ 道 CFA 風格題目）提供實證證據，證明 AI 在可形式化任務上的表現具有格式不變性——支持我們關於訊號侵蝕反映真正知識複製而非評量格式利用的預測。

2. 理論基礎

本文架構整合三條理論脈絡。第一，Spence [13] 的典範訊號模型確立了專業證照僅在取得訊號的成本與能力呈負相關時，方能維繫分離均衡。訊號文獻的一致發現 [12, 14, 5] 指出，當不同類型間的成本差異縮小時，訊號即失去價值——這正是 AI 複製所引發的動態機制。

第二，Becker [4] 區分了一般性人力資本與專屬性人力資本。專業證照主要認證一般性人力資本——跨雇主可攜的可編碼知識——使其在本質上

容易受到 AI 複製的影響。人力資本架構預測一種結構性轉變：一般性認知技能隨 AI 以低成本提供而貶值，而隱性技能則相對升值 [9]。

第三，Autor et al. [3] 的任務導向架構將工作場所活動分類為例行認知、例行操作、非例行分析與非例行互動任務，指出電腦化替代例行任務，同時補充非例行任務。Acemoglu and Restrepo [1] 將此擴展至一般均衡分析。關鍵在於，LLMs 代表一個質性轉變：不同於早期的自動化，它們能夠執行先前被認為抗拒電腦化的某些非例行分析任務，擴展了可自動化的邊界，對證照訊號產生嶄新的意涵。

在實證方面，Callanan et al. [6] 顯示 GPT-4 在 CFA Level I 和 Level II 達到通過水準的表現，而領域適配模型如 BloombergGPT [15] 與 FinDAP 的 Llama-Fin [11] 進一步透過針對性後訓練縮小差距。Patel et al. [10] 證明推理模型（o1、o3）現已通過 CFA 三個等級的考試，分數超過人類考生第 90 百分位數，顯示可形式化能力的邊界正在快速擴展。Galdin and Silbert [8] 將 Spence 訊號理論應用於自由接案勞動市場，指出 AI 生成的作品樣本降低了低能力者的訊號成本，劣化了分離均衡——此動態機制與我們的專業證照場景類似，但處於不同的制度脈絡。我們的架構將其分析從非正式訊號（作品品質）擴展至正式證照（標準化考試），後者的制度利害更高，但底層的經濟邏輯是平行的。這些結果引發我們的核心問題：當 AI 以低成本複製受認證的認知能力時，證照在勞動市場角色的均衡意涵為何？

3. 修正式 Spence 訊號模型：融入 AI 複製

3.1. 模型設定

考慮一個資訊不對稱的勞動市場。存在一連續體的勞工，每位勞工具具有不可觀察的能力類型 $\theta \in \{\theta_L, \theta_H\}$ ，其中 $\theta_H > \theta_L > 0$ 。勞工為高能力者的先驗機率為 $\lambda \in (0, 1)$ 。勞工可以投資取得專業證照 $e \in \{0, 1\}$ （例如 CFA 特許狀），成本為 $c(\theta)$ ，且滿足單交叉性質： $c(\theta_H) < c(\theta_L)$ 。

我們偏離標準模型之處在於引入多維能力空間。令證照 e 認證一向量包含 K 個不同的技能維度：

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

其中每個 s_k 代表證照所欲衡量與傳遞訊號的一項特定認知能力。在 CFA 的脈絡下，這些維度包含陳述性知識（ s_1 ）、演算法運算（ s_2 ）、分析性拆解（ s_3 ）、整合性判斷（ s_4 ）、規範性/倫理推理（ s_5 ）與利害關係人推理（ s_6 ）。

Assumption 1 (雇主估值). 雇主對持證勞工的估值為各技能維度的加權總和：

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

權重 w_k 反映市場對每項技能對勞工生產力貢獻的評估。在前 AI 均衡中，這些權重相對穩定，由生產技術所決定。

3.2. AI 複製成本函數

本模型的關鍵創新在於引入 AI 複製成本函數，捕捉人工智慧系統複製各技能維度的成本。

Definition 1 (AI 複製成本). 對每個技能維度 s_k ，定義 AI 複製成本 $c_{AI}(s_k) \geq 0$ 為 AI 系統產出與具備能力 s_k 之人類勞工等質產出的邊際成本。AI 複製成本向量為：

$$\mathbf{c}_{AI} = (c_{AI}(s_1), c_{AI}(s_2), \dots, c_{AI}(s_K)). \quad (3)$$

Assumption 2 (異質可複製性). AI 複製成本在各技能維度間具有異質性。具體而言，對 CFA 能力空間，存在一劃分 $\{1, \dots, K\} = \mathcal{F} \cup \mathcal{T}$ ，分為可形式化能力 (\mathcal{F}) 與隱性能力 (\mathcal{T})，使得：

$$c_{AI}(s_k) \rightarrow 0 \quad \forall k \in \mathcal{F}, \quad c_{AI}(s_k) \gg 0 \quad \forall k \in \mathcal{T}. \quad (4)$$

此劃分直接源自 Autor et al. [3] 任務導向架構。可形式化能力 (\mathcal{F}) 對應例行認知任務，以及非例行分析任務中可由 LLMs 透過模式複製完成的子集。隱性能力 (\mathcal{T}) 對應非例行互動任務，需要具身化判斷、倫理審議及情境推理，抗拒低成本的 AI 複製。

3.3. AI 可複製性指數

為形式化各技能維度的 AI 脆弱程度，我們定義：

Definition 2 (AI 可複製性). 技能維度 s_k 的 AI 可複製性為：

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

其中 \bar{c}_k 為技能 s_k 的人類取得成本（即人類透過訓練與教育發展此能力所需的投入）。當 $\rho_k \rightarrow 1$ 時，該技能完全可由 AI 複製；當 $\rho_k \rightarrow 0$ 時，該技能仍為人類的比較優勢。

3.4. AI 複製下的訊號價值

在古典 Spence 模型中，證照因能可信地傳達不可觀察的能力而產生訊號價值。AI 複製透過提供替代性的低成本等效產出來源，干擾了此機制。我們將其形式化如下。

Definition 3 (有效訊號價值). 證照 e 在 AI 複製下的有效訊號價值為：

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

其直覺是直接的：技能 s_k 的訊號貢獻被因子 $(1 - \rho_k)$ 所折減，此因子捕捉在扣除 AI 複製後該技能的殘餘稀缺性。當 $\rho_k = 1$ 時，該技能貢獻零訊號價值，因為 AI 以可忽略的成本提供之；當 $\rho_k = 0$ 時，完整的訊號價值得以保留。

3.5. 修正後的雇主信念

在 AI 衝擊下，雇主必須考量持證技能是否仍能區分該勞工與 AI 增強的未持證勞工。雇主觀察 $e = 1$ 的殘餘資訊增益為：

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

當技能 k 的 $\rho_k \rightarrow 1$ 時，項 $(1 - \rho_k) \rightarrow 0$ ，該技能維度不再對雇主的資訊增益有所貢獻。證照在該維度上的資訊量等同於雜訊。

4. 均衡分析：部分訊號崩解

4.1. 前 AI 時代的分離均衡

在標準 Spence [13] 架構中，當以下誘因相容條件成立時，分離均衡存在：

$$\text{高類型：} \quad V(\mathbf{s}) - c(\theta_H) \geq V_0 \quad (8)$$

$$\text{低類型：} \quad V_0 \geq V(\mathbf{s}) - c(\theta_L) \quad (9)$$

其中 V_0 為未持證勞工的市場工資。分離均衡要求 $c(\theta_H) < V(\mathbf{s}) - V_0 < c(\theta_L)$ ：證照的工資溢價對高類型而言超過其成本，但對低類型則不然。

4.2. AI 修正後的均衡條件

在 AI 複製下，雇主對持證技能的支付意願發生變化。證照的有效工資溢價變為：

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

修正後的誘因相容條件為：

$$\text{高類型：} \quad \Pi_{AI} \geq c(\theta_H) \quad (11)$$

$$\text{低類型：} \quad c(\theta_L) \geq \Pi_{AI} \quad (12)$$

Remark 1. 隨著越來越多技能維度的 $\rho_k \rightarrow 1$ ， Π_{AI} 單調遞減。分離均衡僅在 $\Pi_{AI} > c(\theta_H)$ 時得以維繫。當 AI 複製成本在足夠多維度上充分下降時，工資溢價甚至對高類型也降至證照成本以下，分離均衡即崩解。

4.3. 部分訊號崩解定理

我們現在陳述本文的核心理論成果。

Proposition 1 (部分訊號崩解). 令 \mathcal{F} 與 \mathcal{T} 分別表示可形式化與隱性技能維度的集合，且 $|\mathcal{F}| + |\mathcal{T}| = K$ 。假設 AI 複製成本滿足假設 2： $\rho_k \rightarrow 1$ 對所有 $k \in \mathcal{F}$ ，且 $\rho_k \approx 0$ 對所有 $k \in \mathcal{T}$ 。則有效訊號價值收斂至：

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

亦即，證照的訊號價值崩解至僅剩隱性能力的加權總和。訊號侵蝕具有部分性：它消除了可形式化技能的資訊內容，同時保留了隱性技能的訊號價值。

Proof. 由定義 3，有效訊號價值為：

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

在假設 2 下， $\rho_k \rightarrow 1$ 對 $k \in \mathcal{F}$ ，故 $(1 - \rho_k) \rightarrow 0$ 對這些維度成立。同時， $\rho_k \approx 0$ 對 $k \in \mathcal{T}$ ，故 $(1 - \rho_k) \approx 1$ 。取極限：

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

訊號價值由 $\sum_{k=1}^K w_k s_k$ 降低至 $\Sigma_{\mathcal{T}} = \sum_{k \in \mathcal{T}} w_k s_k$ ，代表與可形式化技能在雇主估值中的權重成比例的部分（非全面）崩解。□

Corollary 1 (訊號留存比率). 前 AI 訊號價值中在 AI 衝擊後存續的比例為：

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

若證照課程高度偏重可形式化技能（即 $\sum_{k \in \mathcal{F}} w_k \gg \sum_{k \in \mathcal{T}} w_k$ ），則 $R \rightarrow 0$ ，證照趨近完全訊號失效。

4.4. 臨界點分析

我們現在刻畫分離均衡完全崩解的條件。

Definition 4 (AI 可複製比例). 令 α 表示 AI 可複製性超過門檻 $\bar{\rho}$ （設為 0.9 代表近乎完全複製）的證照能力加權比例：

$$\alpha = \frac{\sum_{k: \rho_k > \bar{\rho}} w_k}{\sum_{k=1}^K w_k} = \sum_{k: \rho_k > \bar{\rho}} w_k \quad (15)$$

Proposition 2 (臨界點). 存在一臨界門檻 $\alpha^* \in (0, 1)$ 使得：

1. 若 $\alpha < \alpha^*$ ，分離均衡得以維繫：證照仍為可信賴的訊號，儘管資訊內容已減損。
2. 若 $\alpha \geq \alpha^*$ ，分離均衡崩解為混同均衡：證照不再能可信地區分能力類型。

臨界門檻由下式決定：

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

Proof. 分離均衡要求 $\Pi_{AI} \geq c(\theta_H)$ 。將 Π_{AI} 寫為：

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

在對訊號最不利的情況下，所有 $\rho_k > \bar{\rho}$ 的維度對總和的貢獻為 $(1 - \rho_k) \approx 0$ ，而 $\rho_k \leq \bar{\rho}$ 的維度則貢獻其全部價值。因此：

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

其中我們使用近似：高度可複製的維度貢獻可忽略，而不可複製的維度按 $(1 - \alpha)$ 的比例貢獻。分離均衡成立若且唯若：

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

求解臨界 α ：

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

當 $\alpha > \alpha^*$ 時，高類型的誘因相容條件被違反，不存在分離均衡。 \square

Corollary 2 (均衡動態). 隨著 AI 能力隨時間提升， α 為弱遞增（新的技能維度跨越 $\bar{\rho}$ 門檻）。若證照機構不調整課程， α 終將超過 α^* ，導致均衡崩解。崩解速度取決於 AI 能力提升的速率以及課程中可形式化技能所佔的比例。

5. CFA 證照之應用

5.1. 將 CFA 能力對應至理論架構

我們現在將理論架構應用於 CFA 考試，將其三級課程對應至我們的六維能力分類體系。表 1 呈現此對應關係，參照 CFA Institute 公布的能力架構以及 Autor et al. [3] 的任務導向分類。

Table 1: CFA 能力分類體系：對應任務架構與 AI 可複製性

技能代碼	能力類型	CFA 等級	Autor (2003) 任務分類	AI ρ_k	訊號留存度
s_1	陳述性知識	I	例行認知	~ 0.95	低
s_2	演算法運算	I–II	例行認知	~ 0.92	低
s_3	分析性拆解	II	非例行分析	~ 0.70	中
s_4	整合性判斷	III	非例行分析	~ 0.45	中高
s_5	倫理推理	II–III	非例行互動	~ 0.30	高
s_6	利害關係人推理	III	非例行互動	~ 0.15	高

圖 1 提供六個能力維度的視覺化比較，對照 CFA 持證專業人士與前沿 AI 系統的能力輪廓。雷達圖使不對稱性一目瞭然：AI 在陳述性知識 (s_1) 與演算法運算 (s_2) 上接近甚至超越人類表現，但在倫理推理 (s_5) 與利害關係人推理 (s_6) 上則大幅落後。

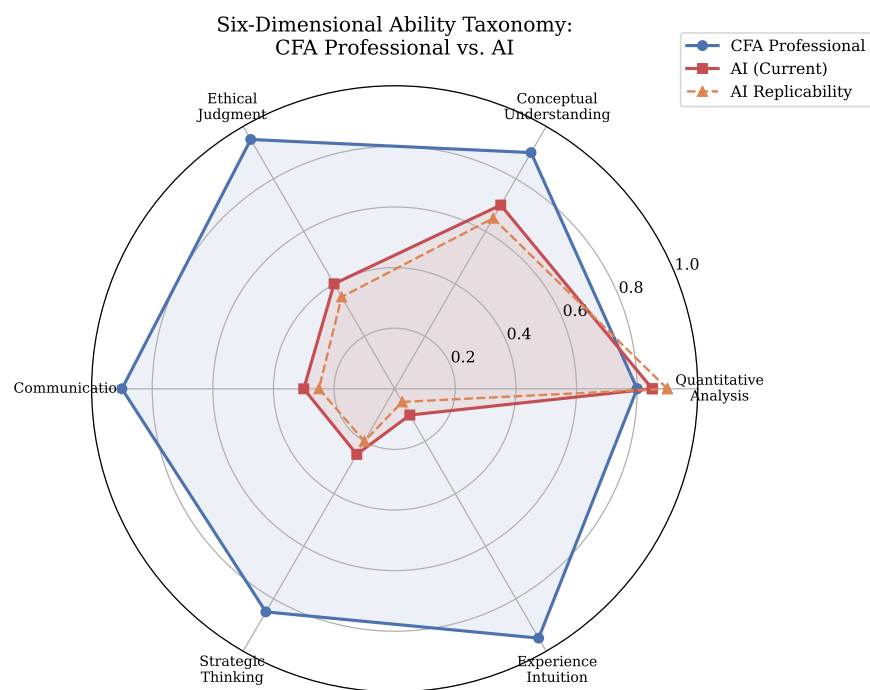


Figure 1: 雷達圖比較 CFA 專業人士與 AI 在六維分類體系中的能力輪廓。每一軸代表一個能力維度 (s_1-s_6)。AI 可複製性在陳述性知識與演算法運算方面接近完全，但在倫理推理與利害關係人推理方面急遽下降，反映部分訊號崩解定理核心的可形式化—隱性劃分。

AI 可複製性數值 (ρ_k) 為基於實證文獻校準的估計值，而非直接估計的參數。對 s_1-s_2 ：Callanan et al. [6] 記錄 GPT-4 在 CFA Level I（主要為陳述性/演算法題項）上達到約 70% 正確率，而 Patel et al. [10] 顯示推理模型達到 97.6%，支持 $\rho \geq 0.90$ 。對 s_3-s_4 ：多步驟分析任務在擾動下顯示大幅退化（我們的姊妹論文記錄了 18.6 百分點的記憶化缺口），將可複製性置於 0.45–0.70 區間。對 s_5-s_6 ：倫理與利害關係人推理題項顯示對抗性脆弱性，但從根本上抗拒低成本複製，支持 $\rho \leq 0.30$ 。從 s_1-s_2 到 s_5-s_6 的急遽下降反映了 Autor et al. [3] 的例行/非例行邊界。關鍵在於，CFA 考試的課程權重配置造成結構性脆弱。表 2 呈現估計的課程權重分布。

Table 2: CFA 課程各能力維度的估計權重分布

能力維度	權重 w_k	類別	ρ_k	$w_k(1 - \rho_k)$
s_1 ：陳述性知識	0.25	\mathcal{F}	0.95	0.013
s_2 ：演算法運算	0.25	\mathcal{F}	0.92	0.020
s_3 ：分析性拆解	0.20	\mathcal{F}/\mathcal{T}	0.70	0.060
s_4 ：整合性判斷	0.15	\mathcal{T}	0.45	0.083
s_5 ：倫理推理	0.10	\mathcal{T}	0.30	0.070
s_6 ：利害關係人推理	0.05	\mathcal{T}	0.15	0.043
合計	1.00			0.288

5.2. 計算訊號留存比率

使用表 2 中的數值，我們計算訊號留存比率（推論 1）：

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

假設所有 k 的 $s_k = 1$ （即正規化能力水準）。這意味著 CFA 證照在當前 AI 能力水準下僅保留約 28.8% 的前 AI 訊號價值。證照歷來向雇主傳達的資訊內容中，超過 70% 現已可由 AI 系統以低成本複製。

圖 2 呈現訊號侵蝕曲線——繪製訊號留存比率 R 作為所有維度加權 AI 可複製性的函數。當前 CFA 的位置 ($R = 0.288$) 已標示，顯示該證照深陷侵蝕區域。曲線突顯侵蝕的非線性本質：一旦可形式化維度（承載最大課程權重者）變為 AI 可複製，訊號價值便急遽下降，之後隨著僅剩隱性能力而趨於平坦。

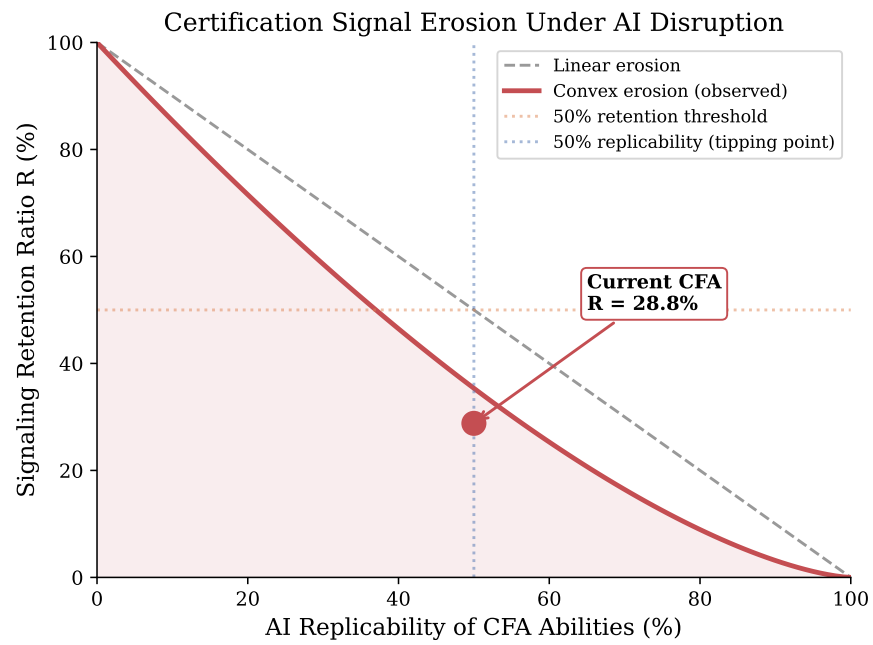


Figure 2: 訊號侵蝕曲線，顯示訊號留存比率 R 隨 AI 可複製性增加的變化。當前 CFA 位置標示於 $R = 28.8\%$ ，表明證照的前 AI 訊號價值已有超過 70% 遭到侵蝕。臨界點 α^* ——超過此點分離均衡崩解為混同——以垂直門檻標示。

5.3. 臨界點評估

使用門檻 $\bar{\rho} = 0.9$ ，目前僅 s_1 與 s_2 符合條件，得 $\alpha = 0.50$ 。若隨 LLMs 持續進步 s_3 跨越 $\bar{\rho} = 0.9$ ， α 將上升至 0.70。表 3 刻畫各情境下的結果。

Table 3: CFA 訊號均衡之臨界點情境

情境	α	相對於 α^*	均衡狀態
當前狀態	0.50	取決於 α^*	分離（弱化）
s_3 跨越 $\bar{\rho}$	0.70	可能 $> \alpha^*$	崩解風險
$s_3 + s_4$ 跨越 $\bar{\rho}$	0.85	幾乎確定 $> \alpha^*$	混同均衡

CFA 當前課程將約 50% 的權重配置於已高度 AI 可複製的能力上，使其處於脆弱區域，在缺乏干預的情況下逼近均衡崩解。

6. 實證證據：選擇題選項作為資訊訊號

我們透過一項控制組實驗提供直接的實證支持，測試選擇題的選項結構是否作為差異性地輔助 AI 表現的資訊訊號。

6.1. 實驗設計

識別策略相當直觀：若選項作為輔助 AI 的資訊訊號（例如透過刪去法），移除選項應降低表現；若 AI 表現反映的是真正的知識複製，選項的有無應無關緊要。我們將 $N = 1,032$ 道來自 CFA-Easy 基準測試 [11] 的 CFA 風格題目施測於 GPT-4o-mini，分為兩種條件：

- 有選項：標準選擇題格式，呈現所有答案選項。
- 無選項：相同的題幹，但移除所有答案選項。模型必須產生自由作答，再由 LLM 評判員對照標準答案評估語意正確性。

此配對設計使得題目內比較成為可能，消除了題目難度變異的干擾因素。關鍵檢定統計量為附帶 Yates 連續性校正的 McNemar 檢定，用於配對名義資料，評估兩種條件下正確/錯誤回答的邊際分布是否有差異。

6.2. 實驗結果

表 4 呈現跨兩個模型世代的選項偏誤實驗彙總結果。

對 GPT-4o-mini 而言，選項偏誤不具統計顯著性 ($p = 0.251$)，不一致配對近乎對稱 ($b = 147$ vs. $c = 127$)，表明為對稱性變異。然而 GPT-5-mini——次世代推理模型——呈現截然不同的模式：選項偏誤擴大至 +9.6 pp 且高度顯著 ($p < 0.001$)，不一致配對強烈不對稱 ($b = 146$ vs. $c = 47$)。推理模型從選項的存在中獲益顯著更多。

Table 4: 選項偏誤實驗結果 ($N = 1,032$)

指標	GPT-4o-mini	GPT-5-mini
有選項正確率	82.6% (852/1,032)	92.8% (958/1,032)
無選項正確率	80.6% (832/1,032)	83.2% (859/1,032)
選項偏誤 (Δ)	+1.9 pp	+9.6 pp
不一致 b (有 \checkmark , 無 \times)	147	146
不一致 c (有 \times , 無 \checkmark)	127	47
χ^2 (Yates 校正)	1.318	49.76
p 值	0.251	< 0.001***

6.3. 對訊號理論的意涵

跨模型比較為格式不變性假說引入了細緻的圖像。對 GPT-4o-mini 而言：

$$\rho_k^{\text{MC}} \approx \rho_k^{\text{free-response}} \quad \text{對 } k \in \mathcal{F} \quad (18)$$

可形式化技能的 AI 可複製性近似格式不變，支持命題 1：訊號侵蝕反映真正的知識複製，而非格式利用。

然而 GPT-5-mini 的顯著選項偏誤 ($p < 0.001$) 揭示格式不變性可能具有世代依賴性。當模型從模式匹配演進至延展式思維鏈推理時，選擇題選項日益成為多路徑審議的收斂錨點。這具有兩個相互競爭的意涵：

1. 樂觀面（對證照而言）：若更強大的模型更依賴格式，則開放式評量格式可能部分恢復對未來 AI 系統的訊號價值。選擇題格式膨脹了 AI 可複製性；轉向自由作答將降低推理模型的衡量 ρ_k 。
2. 悲觀面（對證照而言）：GPT-5-mini 的無選項正確率 (83.2%) 仍超過 GPT-4o-mini 的有選項正確率 (82.6%)。AI 可複製性的絕對水準無論格式如何均持續跨世代上升，意味著格式改革僅延緩而非逆轉訊號侵蝕。

對訊號留存比率 R 的淨效果取決於何種動態佔主導。在悲觀詮釋下——我們的數據更強力支持此詮釋——政策處方仍為內容改革而非格式改革，儘管跨模型證據顯示格式改革可能提供有意義的互補性干預，尤其是對選項偏誤顯著的推理模型。

6.4. 穩健性與適用範圍

本實驗橫跨兩個模型世代 (GPT-4o-mini 與 GPT-5-mini)，題庫以可形式化能力 (s_1 、 s_2) 為主。跨模型在顯著性上的反轉（從 $p = 0.251$ 到 $p < 0.001$ ）證明格式效果並非評量工具的靜態屬性，而是模型與格式互動

的動態屬性。擴展至其他模型家族以及 CFA Level III 針對整合性判斷與倫理推理的論文題，將進一步細化格式不變性的邊界。

7. 意涵與政策建議

7.1. 可檢驗的預測

本模型產生三項關鍵的可檢驗預測：

- P1. 差異性工資溢價侵蝕。**CFA 工資溢價在以可形式化技能為主的職務功能（例如量化分析）中的侵蝕應快於需要隱性技能的職務功能（例如客戶諮詢）。溢價變動 $|\Delta w_j|$ 應與所需技能的平均 AI 可複製性 $\bar{\rho}_j$ 呈正相關。
- P2. 雇主行為轉變。**理性雇主應以針對隱性能力的方法（行為面試、案例模擬）補充證照篩選，且採用率應隨證照核心內容的 AI 可複製性增加而上升。
- P3. 格式不變性（部分獲支持）。**可形式化技能的 AI 可複製性應近似格式不變。GPT-4o-mini 支持此預測（+1.9 pp, $p = 0.251$ ），但 GPT-5-mini 顯示顯著的格式依賴（+9.6 pp, $p < 0.001$ ），暗示格式不變性可能對推理模型而言會侵蝕，但絕對可複製性持續上升。

7.2. 證照設計的政策意涵

命題 1 與 2 為證照機構衍生三項處方。

課程再平衡。降低 s_1 （陳述性知識）與 s_2 （演算法運算）的權重，它們高度 AI 可複製（ $\rho > 0.9$ ）且對殘餘訊號價值的貢獻微乎其微。增加 s_5 （倫理推理）與 s_6 （利害關係人推理）的權重，它們的 AI 可複製性低（ $\rho < 0.3$ ）且代表留存訊號價值的主要來源。CFA Level III 論文格式的 IPS 題目方向正確但不足，因整體課程仍偏重可形式化技能。

格式創新搭配內容改革。我們的實證證據顯示，單獨的格式變更並不足夠：移除選擇題選項並未降低 AI 表現（ $p = 0.251$ ）。格式改革必須針對隱性技能維度（ s_4-s_6 ），透過互動式案例模擬、無唯一正確答案的倫理困境審議，以及候選人必須批判性評估並推翻 AI 建議的 AI 增強評量。

動態校準。隨著 AI 能力擴展可複製邊界（推論 2），證照機構應建立明確連結 AI 能力基準的定期檢視機制 [6, 11]。

8. 討論

8.1. 擴展與一般化

本模型補充了 Acemoglu and Restrepo [1] 的一般均衡方法，聚焦於 AI 對特定市場制度的資訊性影響。關鍵洞見在於：即使 AI 未消滅工作，它仍可能削弱組織勞動市場的篩選制度——社會成本是錯配而非替代。

我們的可形式化—隱性區分連結至「波蘭尼悖論」[2]：LLMs 複製隱性推理的表層結構，卻缺乏深層結構（規範承諾、法律責任、聲譽利害相關）。此架構可推廣至任何作為勞動市場訊號運作的證照：偏重可形式化內容的證照（FRM、基礎精算考試）應經歷更快的侵蝕，而強調隱性能力的證照（醫學委員會臨床考試、口試制律師考試）則應侵蝕較慢。

8.2. 研究限制

應承認若干限制。第一，我們的實證證據完全來自 A5 選項偏誤實驗；訊號模型本身依賴對 AI 可複製性 (ρ_k) 的校準假設，這些假設雖受文獻啟發，但並非直接估計而得。將所有 ρ_k 值擾動 ± 0.10 會使訊號留存比率 R 移至約 $[0.19, 0.39]$ ——此範圍不改變實質性侵蝕的定性結論，但確實影響 28.8% 估計值的精確度。第二，可形式化—隱性的二分法是一種簡化：實務上，金融能力存在於連續光譜而非整齊地落入兩個類別。例如，分析性拆解 (s_3) 結合了可形式化與隱性要素，其分類可能隨 AI 能力演進而轉移。第三，我們缺乏雇主行為的直接證據——雇主是否實際上在因應 AI 複製 CFA 認證能力而調整其聘僱實務。第 7 節中的可檢驗預測源自模型但在實證上尚未受到檢驗。第四，跨世代證據僅橫跨單一供應商 (OpenAI) 的兩個模型；擴展至其他模型家族將強化實證基礎。第五，GPT-5-mini 的 A5 無選項正確率 (83.2%) 包含對 58 個空白回應視為錯誤的校正。使用未校正值 (86.3%) 將產生 $R \approx 0.31$ 而非 0.288——結論方向不變，但應注意此資料品質問題的敏感性。

9. 結論

我們發展一個修正式 Spence 訊號模型，顯示 AI 驅動的訊號侵蝕具有部分性與選擇性：證照在可形式化能力上失去資訊內容，同時保留隱性能力的訊號價值。一項跨世代選項偏誤實驗 ($N = 1,032$) 揭示格式不變性具有世代依賴性——對 GPT-4o-mini 不顯著 ($p = 0.251$) 但對 GPT-5-mini 高度顯著 ($p < 0.001$)——然而 AI 可複製性的絕對水準無論格式如何均持續上升，支持內容改革優先於僅格式改革的處方。

應用於 CFA，約 50% 的課程權重已高度 AI 可複製，且該證照僅保留約 29% 的前 AI 訊號價值。跨模型證據引入一個時間維度：每個模型世代同時提升基準可複製性並可能改變格式敏感性的格局，需要持續校準 AI 可複製性指數 ρ_k 。政策處方明確：證照機構必須朝倫理推理、整合性判斷與利害關係人審議重新平衡評量內容——並必須建立連結 AI 能力基準的動態檢視機制，因為訊號格局隨每個模型世代而遷移。

更廣泛而言，我們的架構證明 AI 不僅重塑生產，更重塑組織經濟活動的制度。跨世代證據強調此重塑並非一次性事件，而是持續性過程，需要以 AI 進步的步調進行制度性調適。

資料可得性

選項偏誤實驗所使用的 CFA-Easy 基準測試($N = 1,032$)可透過 FinDAP 架構取得 [11]。實驗程式碼與完整結果 (JSON) 可向通訊作者索取。

利益衝突聲明

作者聲明並無已知的競爭性財務利益或個人關係可能影響本文所報導之研究。

CRedit 作者貢獻

Wei-Lun Cheng：概念化、方法論、形式分析、軟體、研究調查、撰寫——初稿。**Daniel Wei-Chung Miao**：指導、撰寫——審閱與編修。**Guang-Di Chang**：指導、撰寫——審閱與編修。

致謝

計算資源由國立臺灣科技大學 (NTUST) 提供。

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., Daberijs, V., Dredze, M., Gehrmann, S., Kambadur, P., Rosenberg, D., Mann, G., 2023. BloombergGPT: A large language model for finance. *arXiv preprint arXiv:2303.17564*.