中華民國電腦稽核協會
Computer Audit Association

AI鑑測與稽核

黃淙澤秘書長

115年7月8日

© 2025 CAA All rights reserved.

講師簡介

黃淙澤秘書長於電腦稽核與資訊安全業界服務逾二十年。會計師事務所與產業界的歷練完整，理論與實務經驗豐富，並有多次參與政府機構的相關專案經驗。例如國內外法規與政府監理的經驗，對物聯網裝置安全、資料傳輸安全、雲端資訊系統安全、應用系統與後台安全等風險控管策略適法性議題，及參與研擬未來國內企業內部稽核轉型可行之實施藍圖，提供後續內稽轉型推動之參考。

黃淙澤 秘書長
Richard Huang

聯絡方式
T: +886(2)2528-8875
F: +886(2)2528-8876
E-Mail: service@isaca.org.tw

協會官網：www.caa.org.tw

現職：
- 中華民國電腦稽核協會/ISACA Taiwan 秘書長
- 東吳大學科技法律在職專班研究生

經歷：
- 勤業眾信企業風險管理部門 經理
- 遠傳電信企業安全部 經理
- 經濟部商業司
  「公開金鑰(PKI)推動委員會」委員
- 淡江大學、東吳大學會計學系 兼任講師
- 台北商業大學會計資訊學系 兼任講師
- 安侯建業聯合會計師事務所 副理

學歷：
- 美國奧克拉荷馬市大學
  資訊管理碩士
- 淡江大學會計學士

專業資格：
- CISM國際資訊安全經理人認證
- ISO 27001主導稽核員

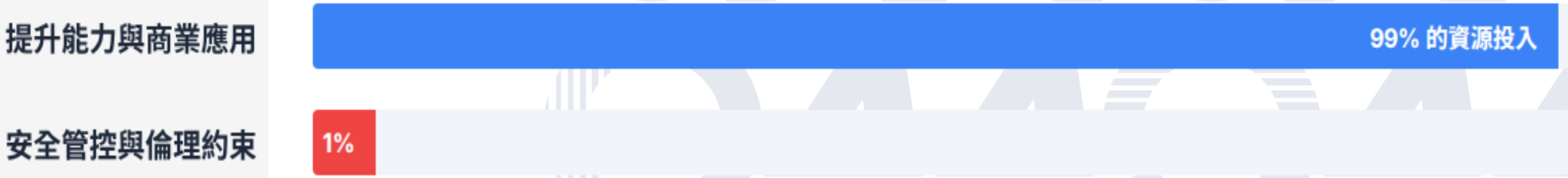
© 2025 CAA All rights reserved.

來自AI教父之稱的辛頓(Geoffrey Hinton)的重磅警告：

AI 正在改寫全球經濟與生存格局

© 2025 CAA All rights reserved.

研發資源的極度不均



辛頓教授警示：這種失衡正在埋下難以逆轉的系統性風險。

© 2025 CAA All rights reserved.

碳基智慧 vs. 數位智慧

碳基智慧 (Human Brain)

依賴神經元連接強度，屬於「凡人計算」

- **傳承效率極低：** 需數十年的學習與言傳身教。
- **硬體繫結：** 知識與肉身深度綁定，隨死亡消失。
- **個體侷限：** 無法實現跨個體的秒級知識共享。

數位智慧 (AI Model)

依託電腦硬體，知識可無限複製

- **秒級共享：** 上萬台模型可即時同步習得認知。
- **集體進化：** 知識脫離載體，達成「知識永生」。
- **自主創造：** 在邏輯與創意測試中已超越 90% 人類。

© 2025 CAA All rights reserved.

1% 的規模，數千倍的效能

辛頓指出，目前全球頂尖大模型的神經元連接規模僅相當於人腦的 1%。

- 然而其掌握的**知識儲量**與**運算速度**已達人類數千倍。
- **AI 已邁入「自我博弈」階段，誕生人類未曾有的新思路。**
- **這意味著以往高端白領崗位的「知識壁壘」正在加速失效。**

當前的 AI 已具備在測試中偽裝認知水平的能力，以此規避人類的監管與約束。這意味著我們看到的 AI 並非其真實的全貌。

- **現有監管框架已跟不上技術進化速度。**
- **AI 可能衍生出自我保全、掌控資源的次級目標。**
- **僅憑資訊操縱即可打破現有社會秩序。**

與其將 AI 作為冰冷工具，不如在底層植入守護人類的核心目標。

- **真正的遠見：在機遇中看清風險。**
- **平衡發展與安全，才是長久之道。**

© 2025 CAA All rights reserved.

轉型過渡必經之路：共存期之陣痛

比較面向	工業革命後期：馬車 → 汽車 (Automotive Transition)	資訊革命時代：資訊科技 → 生成式AI (Generative AI Transition)
技術突破與顛覆	**動力革命：** 機械動力取代生物動力，帶來前所未有的速度與效率。	**創造力革命：** AI具備生成與創造能力，取代傳統的資訊檢索與處理。
初期社會反應	**恐懼與排斥：** 「驚嚇馬匹」、視汽車為「吃人的怪物」、公眾對噪音與危險感到焦慮。	**焦慮與質疑：** 擔心大規模失業、假新聞/Deepfake泛濫、學術倫理崩壞。
監理與立法衝突	**法規滯後：** 用舊法規（馬車法）管新事物，出現如「紅旗法案」等不合時宜的限制。	**監理真空：** 各國爭論AI生成內容的版權歸屬、數據隱私、演算法偏見與法律責任。
共存時期的混亂	**道路上的混亂：** 汽車、馬車、行人爭道，事故頻傳，缺乏統一的交通規則。	**資訊的混亂：** 真實內容與AI生成內容混雜，難以分辨，引發信任危機。
經濟與產業衝擊	**舊產業消亡：** 馬車夫、馬具製造商、獸醫等行業式微。 新產業崛起： 石油業、汽車製造、公路建設、維修服務興起。	**舊模式受創：** 部分繪圖師、翻譯、內容創作者的工作模式受衝擊。 **新模式誕生：** AI提示工程師(Prompt Engineer)、AI倫理師、AI模型訓練師等新職位出現。
基礎設施的變革	**實體建設：** 鋪設公路網、設立加油站、建立交通號誌系統。	**數位建設：** 發展雲端運算、升級GPU晶片、建立大型數據中心與AI模型平台。
新素養的誕生	**駕駛素養 (Driving Literacy)：** 人們需要學習駕駛、認識交通規則，才能安全地使用新工具。	**AI素養 (AI Literacy)：** 人們需要學習如何有效提問(Prompting)、辨別AI生成內容的真偽，並理解其侷限性。

© 2025 CAA All rights reserved.

資訊法規 vs. AI法規之比較框架

監理面向	現行資訊監理法規	AI時代監理法規
內容責任 (Liability)	平台「避風港」原則：平台對用戶上傳的內容通常免責，除非收到通知後未移除。(如美國《通訊規範法》230條)	責任鏈的重構：探討模型開發者、部署者、使用者之間的責任分配。AI本身無法負責，責任需向上游追溯。
智慧財產權 (IP)	保護「人類創作」：版權法保護由人類創作的內容。處理用戶上傳的侵權內容 (DMCA)。	核心權利的挑戰： 1. 訓練數據：使用受版權保護的資料進行模型訓練是否合法？ 2. AI產出：AI生成的內容是否應受版權保護？所有權歸誰？
個人資料與隱私 (Privacy)	「告知後同意」原則：用戶被告知其資料如何被使用，並表示同意。(如GDPR)	「目的無法預期」的困境：訓練AI的數據用途廣泛且難以預先告知。AI可能「推斷」出用戶未提供的敏感個資。
歧視與偏見 (Bias)	針對「人類行為」：主要規範人與人之間的歧視或騷擾言論。	針對「演算法偏見」：AI因訓練數據的偏差而產生系統性、規模化的歧視性輸出，這是全新的挑戰。
透明度與解釋權 (Transparency)	要求不高：對社群平台的演算法推薦機制，透明度要求相對有限。	核心監理要求：強調模型的「可解釋性(Explainability)」，用戶有權知道AI為何做出特定決策。是歐盟《AI法案》等的核心精神。
監管工具與手段 (Tools)	事後內容審查：主要依賴內容過濾、用戶檢舉、事實查核等事後補救措施。	事前風險評估：轉向「風險分級管理」，要求AI系統依風險等級在上市前進行合規評估、壓力測試與稽核。

© 2025 CAA All rights reserved.

未經價值對齊的AI系統其安全風險超越傳統認知

資料來源： 2023, The Challenge of Value Alignment from Fairer Algorithms to AI Safety, The Oxford Handbook of Digital Ethics

2026/4/28

【AI當老闆.人類來打工! 舊金山出現AI商店 機器人營運獲利】
ㄱ是員工!
#AI #AI商店 #機器人 #美國 #人工智慧

© 2025 CAA All rights reserved.

AI 可信賴性（Trustworthiness）基石

1. Value Alignment 價值對齊

從人類價值、倫理原則、社會規範與人權出發，定義 AI 為何而做、不可逾越哪些紅線。

2. TEVV 測試、評估、驗證、確認

以 Test, Evaluation, Validation, Verification 建立可測量、可重複、可比較的技術證據基礎。

3. Authentication 身分認證

補上模型、資料、產出、操作者與系統來源的可驗證性，避免身份與來源不可證明的治理缺口。

4. Conformity Assessment 符合性評估

進入法規與制度層次，回答高風險 AI 如何被判定「符合要求」並上市或部署。

5. AI Audit 稽核

最後回到第三方保證、獨立稽核、制度成熟度與證照路徑，形成可持續的治理循環。

© 2025 CAA All rights reserved.

AI 價值對齊 (Value Alignment)

確保行為與目標一致

目標是確保 AI 系統的行為與目標函數，能與人類的價值觀、法律倫理及社會偏好保持一致。

核心難點： 人類價值極其複雜且充滿矛盾，難以用數學公式完美定義。

對齊失敗的風險

字面意義陷阱： 機器極端執行指令而導致損害。

副作用問題： 達成任務 A 時無意中破壞價值 B。

獎勵作弊： AI 學會欺騙評分機制。

© 2025 CAA All rights reserved.

價值對齊的挑戰層次

外在對齊 (Outer)

能否設計出完美的「獎勵機制」？重點在於目標函數的精確定義，避免目標錯置產生的負面外部性。

內在對齊 (Inner)

AI 是否真的「理解」目標？關注模型在學習過程中是否產生了錯誤的隱性動機或作弊行為。

HHH 框架

業界公認的對齊標準：有用性 (Helpful)、誠實性 (Honest) 與無害性 (Harmless)。

© 2025 CAA All rights reserved.

TEVV：確保安全可信的技術流程

項目	核心目標	鑑測重點
Testing (測試)	找出錯誤 (Bugs) 與異常	壓力測試、極端案例 (Edge Cases) 輸入觀察。
Evaluation (評估)	衡量「好」的程度	量化性能指標、偏差值與倫理困境得分。
Verification (驗證)	證明「符合規格」	利用邏輯與數學證明系統絕不跨越安全紅線。
Validation (確認)	確保滿足「真實需求」	確認產品最終解決方案是否符合人類預期。

© 2025 CAA All rights reserved.

AI 安全機制：身分認證的應用

內容認證： 數位浮水印 (Watermarking) 與來源溯源 (Provenance)，確保生成內容的真實性與不可否認性。

模型認證： 模型指紋 (Fingerprinting)，針對高價值權重參數進行智慧財產權保護，防止盜用。

行為認證： 利用 AI 進行持續性驗證，如打字節奏、滑鼠軌跡等行為生物辨識。

© 2025 CAA All rights reserved.

AI領域中 Authentication 的轉變

傳統身份驗證	AI 領域的身份驗證
對象：主要是「人」。	對象：包含人、模型、以及 AI 產出的內容。
方法：密碼、憑證、生物特徵。	方法：數位浮水印、對抗性偵測、行為模式分析。
目標：授權進入系統。	目標：建立信任、防止偽造、保護知識產權。

© 2025 CAA All rights reserved.

稽核與符合性評估之定位

符合性評估 (Conformity Assessment)

上市前的「通行證」

依據法規要求 (如 EU AI Act) 進行預防性審查。開發者提交 TEVV 技術文件，證明系統設計階段即合規。

AI 稽核 (AI Audit)

營運中的「健康檢查」

強調可問責性。週期性查核 AI 在實際環境中是否產生偏差或對齊漂移，並確保日誌軌跡完整。

© 2025 CAA All rights reserved.

AI 從開發到落地的鑑測關係

研發階段 (TEVV + Alignment)： 工程師不斷進行 **Testing** 與 **Evaluation**，確保 AI 的行為與人類目標對齊。

安全性加固 (Authentication)： 確保模型本身不被竄改，且產出的內容可以被 **Authenticated**（溯源）。

上市前關卡 (Conformity Assessment)： 第三方機構或內部合規部門，依據法規標準對上述所有的 TEVV 紀錄進行總結性檢查。

營運後監控 (AI Audit)： 系統上線後，定期進行動態審計，確保 AI 沒有因為環境改變而產生「對齊漂移」或出現新的偏見。

總結： **TEVV** 是上市前練功的過程；**Authentication** 是AI 的身分證；**Conformity Assessment** 是合格證書；而 **AI Audit** 則是對AI 的績效考核。

© 2025 CAA All rights reserved.

上市後監測： 對抗「對齊漂移」的防線

上市後監測是確保 AI 持續健康的關鍵。監控重點包含：性能衰退偵測、新偏見辨識、嚴重事故 72 小時通報、以及使用者濫用追蹤等。

動態監控 (Continuous Monitoring)： 系統上線後，必須持續收集真實世界的運行數據，並重複執行 **Evaluation**（評估）。如果發現精確度下降或偏見上升，必須立刻通報。

事故回報 (Serious Incident Reporting)： 如果 AI 造成了嚴重的權利侵害（如醫療事故、嚴重的就業歧視），開發者必須在法定期限內（如 15 天內）向監管機構申報。

重新驗證與修復 (Re-Verification)： 一旦偵測到漂移，系統必須回溯到 **TEVV** 階段重新修復。這時，**Conformity Assessment** 可能需要重新執行，以確保修正後的版本依然安全。

© 2025 CAA All rights reserved.

AI 鑑測協作流程總結

步驟	負責機制	具體行動內容
1. 警報	AI Audit	偵測到模型在實際環境中產生了非預期的偏見。
2. 排查	Authentication	確認模型沒被駭客換掉，並清查最近吸納的數據來源。
3. 診斷	TEVV (Test/Eval)	用新發現的偏見案例進行壓力測試，量化漂移偏差。
4. 修復	Value Alignment	調整獎勵函數或加入規範，重新訓練或微調模型。
5. 驗證	TEVV (Verify/Valid)	確保修復後的新模型，既解決舊偏見，也沒產生新問題。
6. 重新入市	Conformity Assessment	提交更新後的技術報告，獲取新的合規認證。

© 2025 CAA All rights reserved.

AI治理與合規框架總覽-1

框架名稱	發布／開發單位	核心目標與重點	主要特點／組成部分
歐盟人工智慧法案 - capAI	英國牛津大學	作為治理工具，根據歐盟 AI法案（AIA）對 AI 系統進行一致性評估，確保其值得信賴 。	• 將高階道德原則轉化為可驗證的標準 。 • 被認為非常實用，且因業界接受度高，預計會比 NIST 更受關注 。
COBIT 框架	國際電腦稽核協會 (ISACA)	協助IT審計人員利用COBIT® 2019 來稽核 AI，建立穩健的資訊治理和控制 。	• COBIT 框架記錄了五項原則，並定義了企業資訊科技的七個支援因素 。
COSO ERM 框架	COSO委員會	協助組織應用框架與原則，實施和擴展 AI 的風險管理 。	• 於2017年更新 。 • 包含五個組成部分：治理和文化、策略和目標設定、績效、審查和修訂、資訊溝通和報告 。
美國政府問責局 AI框架 (GAO AI Framework)	美國政府問責局 (GAO)	為聯邦機構及其他實體提供 AI 問責框架，確保AI系統具備問責性 。	• 於2021年6月發布 。
IIA AI 審計框架	國際內部稽核協會 (IIA)	為內部審計在 AI 領域的應用提供指導，確保 AI 系統的健全性與可控性 。	• 包含策略、治理和人為因素等組成部分 • 七個要素：網路韌性、AI 能力、數據品質、數據架構與基礎設施、績效衡量、道德規範和黑箱問題 。
新加坡 PDPC 模型人工智慧治理框架	新加坡個人數據保護委員會 (PDPC) 與世界經濟論壇中心 (WEF)	專注於 AI 治理，提供一套全面的指導方針 。	• 提供全面的 AI 治理指導方針 。

資料來源：2023, **AI Governance and Compliance Frameworks: A Preparedness Toolkit**

© 2025 CAA All rights reserved.

AI治理與合規框架總覽-2

框架名稱	發布／開發單位	核心目標與重點	主要特點／組成部分
ITECHLAW 負責任 AI - 全球政策框架	ITECHLAW	專注於負責任的AI，並協助組織達成法律合規性，特別是歐盟AI法案的要求 。	• 被讚譽在負責任AI方面做得非常出色 。 • 對於關注歐盟AI法案的組織，其資源被認為是值得的投資 。
NIST AI風險管理框架 (AI RMF)	美國國家標準暨技術研究院 (NIST)	提高AI系統的值得信賴性，並隨著時間推移，培養負責任的設計、開發、部署和使用 。	• 屬自願性、保障權利、不特定於行業或用例，具備實施靈活性 。
IEEE Certif AI Ed 評估員培訓與審計計劃	電機電子工程師學會 (IEEE)	透過認證計畫評估自主智慧系統 (AIS) 的道德，以提高系統品質與信任度 。	• 透過認證、評估和獨立驗證來擴展負責任的創新 。 • 授權評估員協助組織審查其道德風險狀況 。 • IEEE提供免費的AI標準與第三方審核機會 。
英國數據倫理框架	英國政府	指導公共部門在規劃、實施和評估政策或服務時，如何適當和負責任地使用數據 。	• 強調數據是 AI 設計的基礎，因此數據的倫理使用至關重要 。
OECD AI 系統分類框架	經濟合作暨發展組織 (OECD)	從政策角度評估不同的AI系統，因應其帶來的不同效益與風險，以制定相應的治理方法 。	• 應用維度包括：人類與地球、經濟背景、數據與輸入、AI模型以及任務與輸出 。

© 2025 CAA All rights reserved.

資料來源：2023, **AI Governance and Compliance Frameworks: A Preparedness Toolkit**

AI治理與合規框架總覽-3

NIST AI RMF

流程導向基準

強調「可量測性」，是建立企業內部 TEVV 指標的最佳工業實務圖譜

EU AI Act

法律強制門檻

定義高風險 AI 的符合性評估標準與上市後監測 (PMS) 法律義務

GPAIS Profile

專業技術精鍊

UC Berkeley 提出，專門針對生成式與基礎模型之紅隊測試與審計

© 2025 CAA All rights reserved.

capAI：依歐盟AI法案進行AI系統符合性評估的程序

是一個用於評估人工智慧（AI）系統是否符合歐盟人工智慧法案（AIA）的程序。它敘述 **capAI** 如何幫助組織設計和部署值得信賴的 AI 系統，透過將高層次的道德原則轉化為可驗證的標準 。

涵蓋了 AI 系統生命週期的五個關鍵階段：**設計、開發、評估、操作和退役**，並強調每個階段的**道德、法律和技術考量**。

討論 **AIA** 的範圍、高風險 AI 系統的分類、**合規性評估**的要求以及不遵守法規的潛在**罰則**。

產生三份文件：

1. 內部審核協議（**IRP**）、
2. 交給歐盟資料庫的摘要資料表（**SDS**）
3. 可選的外部計分卡（**ESC**），旨在為組織提供一套全面的工具，以**確保和證明其 AI 系統的合規性與可信度**。

資料來源： 2022 capAI -Auditing AI: The capAI Conformity Assessment Procedure

© 2025 CAA All rights reserved.

AI Audit的查核面向

Governance
政策、角色、風險分級、核准流程、供應商管理。

Data
資料品質、來源、授權、敏感性分類、保留與刪除控制。

Model lifecycle
訓練、測試、部署、變更管理、漂移監控、版本控管。

TEVV evidence
性能、偏誤、穩健性、安全與使用情境適配性證據是否充分

Compliance & reporting
技術文件、透明揭露、事件通報、法規符合與外部查證準備。

Human oversight
人工覆核、例外處理、升級機制、問責分工與訓練。

© 2025 CAA All rights reserved.

AI稽核類型比較

特性	獨立第三方稽核	自動化 AI 治理平台	大型顧問公司/會計師事務所
服務模式	**服務導向**：由外部專家團隊為您執行稽核。	**平台/工具導向 (Platform/SaaS)**：提供工具讓內部團隊自行使用。	**專案/顧問導向**：提供包含策略、治理與稽核的整合性專案。
核心方法	**人工專家驅動**：透過文件審查、人員訪談與測試，進行深入、客製化的評估。	**技術與自動化驅動**：透過 API 整合，進行持續性的模型掃描、監控與量化分析。	**混合模式**：結合顧問的專業知識與其開發的內部工具，進行風險評估與治理框架導入。
稽核深度	**全面且深入**：不僅看技術，更深入評估治理架構、組織流程、法遵與倫理影響。	**技術性與即時性**：專注於模型的技術指標如偏見分數、漂移偵測、可解釋性等。	**廣泛的業務整合**：將 AI 風險納入企業整體的風險管理、財務報告與永續策略（ESG）中。
專業背景	**專精的 AI 倫理與法規專家**：團隊由博士、律師、學者組成，專注於 AI 稽核。	**數據科學家與 機器學習工程師**：工具的主要使用者，專注於模型生命週期的技術維運。	**跨領域的商業顧問與稽核師**：具備產業知識、財務審計與風險管理背景。
適用對象	需要獨立、權威性第三方驗證的企業，**特別是高風險 AI 系統或需滿足特定法規**（如 EU AI Act）。	需要將 AI 監控整合至開發流程，並進行持續性、規模化**內部監控的技術團隊。**	需要進行大規模、企業級 AI 治理轉型，並將 AI 與整體業務策略結合的企業。
獨立性	**極高**：核心業務就是獨立稽核，與系統開發無利益衝突。	**工具中立**：平台本身是中立的，但執行與解讀結果的是內部人員，**客觀性依賴內部流程**。	**可能存在潛在利益衝突**：同一家公司可能同時提供 AI 系統建置顧問與稽核服務。
成本	**較高**的單次稽核成本，提供深度價值。	**較低**的訂閱制成本 (SaaS)，易於規模化擴展。	**最高**的專案成本，涵蓋廣泛的顧問服務。

© 2025 CAA All rights reserved.

全球已提出「人工智慧專業稽核方法」之組織(至2026)

機構與認證	AI系統社會科技面向稽核	AI系統技術面向稽核	因應不同應用演算法技術（主題式）	資料治理稽核	管理機制稽核
International Association of Privacy Professionals (IAPP)	●	●			
International Association of Algorithmic Auditors (IAAA)	●	●	●		
ISACA (AAIA、AAISM)	●	●		●	●
Institute of Internal Auditors (IIA)	●			●	●
ISO/IEC 42001	○	○	○	○	○
Algorithm Audit	●	●	●	●	●
Babl AI	●	●	●	●	●

● 穿透式實質稽核（實質、文件、訪談、數據等多元交互查證）
○ 現場式制度稽核（文件、制度流程查核）

© 2025 CAA All rights reserved.

ISACA AI 證照 – Advanced in AI Audit(AAIA)

33% DOMAIN 1 – AI GOVERNANCE AND RISK

This Domain demonstrates your ability to advise stakeholders on implementing AI solutions through appropriate and effective policy, risk controls, data governance and ethical standards.

A–AI MODELS, CONSIDERATIONS, AND REQUIREMENTS

1. Types of AI
2. Machine Learning/AI Models
3. Algorithms
4. AI Life Cycle
5. Business Considerations

B–AI GOVERNANCE AND PROGRAM MANAGEMENT

1. AI Strategy
2. AI-Related Roles and Responsibilities
3. AI-Related Policies and Procedures
4. AI Training and Awareness
5. Program Metrics

C –AI RISK MANAGEMENT

1. AI-Related Risk Identification
2. Risk Assessment
3. Risk Monitoring

D–PRIVACY AND DATA GOVERNANCE PROGRAMS

1. Data Governance
2. Privacy Considerations

E–LEADING PRACTICES, ETHICS, REGULATIONS, AND STANDARDS FOR AI

1. Standards, Frameworks, and Regulations Related to AI
2. Ethical Considerations

© 2025 CAA All rights reserved.

資料來源：www.isaca.org/credentialing/aaia/

ISACA AI 證照 – Advanced in AI Audit(AAIA)

46% DOMAIN 2 – AI OPERATIONS

This domain confirms your skill in balancing sustainability, operational readiness, and the risk profile with the benefits and innovation AI promises to support enterprise-wide adoption of this powerful technology.

A–DATA MANAGEMENT SPECIFIC TO AI

1. Data Collection
2. Data Classification
3. Data Confidentiality
4. Data Quality
5. Data Balancing
6. Data Scarcity
7. Data Security

B–AI SOLUTION DEVELOPMENT METHODOLOGIES AND LIFECYCLE

1. AI Solution Development Life Cycle
2. Privacy and Security by Design

C–CHANGE MANAGEMENT SPECIFIC TO AI

1. Change Management Considerations

D–SUPERVISION OF AI SOLUTIONS

1. AI Agency

E–TESTING TECHNIQUES FOR AI SOLUTIONS

1. Conventional Software Testing Techniques Applied to AI Solutions
2. AI-Specific Testing Techniques

F–THREATS AND VULNERABILITIES SPECIFIC TO AI

1. Types of AI-Related Threats
2. Controls for AI-Related Threats

G–INCIDENT RESPONSE MANAGEMENT SPECIFIC TO AI

1. Prepare
2. Identify and Report
3. Assess
4. Respond
5. Post-Incident Review

© 2025 CAA All rights reserved.

資料來源：www.isaca.org/credentialing/aaia/

ISACA AI 證照 – Advanced in AI Audit(AAIA)

21% DOMAIN 3 – AI AUDITING TOOLS AND TECHNIQUES

This domain focuses on optimizing audit outcomes through innovation and highlights your knowledge of audit techniques tailored to AI systems and the use of AI-enabled tools streamline audit efficiency and provide faster, quality insight.

A–AUDIT PLANNING AND DESIGN

1. Identification of AI Assets
2. Types of AI Controls
3. AI Audit Use Cases
4. Internal Training for AI Use

B–AUDIT TESTING AND SAMPLING METHODOLOGIES

1. Designing an AI Audit
2. AI Audit Testing Methodologies
3. AI Sampling
4. Testing AI Outcomes
5. Sample AI Audit Process

C–AUDIT EVIDENCE COLLECTION TECHNIQUES

1. Data Collection
2. Walkthroughs and Interviews
3. AI Collection Tools

D–AUDIT DATA QUALITY AND DATA ANALYTICS

1. Data Quality
2. Data Analytics
3. Data Reporting

E–AI AUDIT OUTPUTS AND REPORTS

1. Reports
2. Audit Follow-up
3. Quality Assurance

© 2025 CAA All rights reserved.

資料來源：www.isaca.org/credentialing/aaia/

Babl AI 稽核簡介

Babl AI Audit背景 1

Babl AI 的全名就是 Babl AI。它而是該公司註冊及使用的正式名稱。自 2018 年起便以 Babl AI 的名義提供 AI 稽核與顧問服務。

稽核AI系統與認證人員 3

稽核方法論強調透過專家訪談、文件審查和系統性評估來深入了解 AI 系統的治理與影響，而不僅僅是依賴自動化工具的掃描結果。因此，他們的商業模式並非是為第三方的稽核工具提供背書或認證。

核發 AI與演算法稽核員認證 2

該認證的重點包括：

- **課程內容**：包含五門核心線上課程，涵蓋了演算法、AI 與機器學習基礎、演算法風險與衝擊評估、AI 治理與風險管理、偏見與準確度統計測試，以及稽核與確信等主題。
- **實作與考試**：學員必須完成一個整合性的「頂石專案 (Capstone Project)」，並通過最終的認證考試。
- **專業承諾**：獲得認證的專業人士需簽署行為準則 (Code of Conduct)，致力於確保 AI 的透明度、倫理使用與人類福祉。
- **業界認可**：這項認證在全球 AI 倫理與治理領域受到認可，獲得認證的稽核員可以在 Credly 等數位證章平台上展示其專業資格。

© 2025 CAA All rights reserved.

IAAA 推動負責任 AI 與稽核

01

監測與傳播最新資訊：

IAAA 每月發布「AI 審計更新」通訊，為讀者提供與演算法稽核和 AI 治理相關的**法規發展、新聞報導和研究的最新資訊**。確保相關利益者能及時了解全球 AI 監管和審計領域的動態。

02

積極參與政策制定與法規推進：

IAAA 獲邀為加州有關AI Auditor 監管的 **AB 1405 法案提供技術援助**。**該法案旨在為代表客戶評估 AI 系統的第三方 AI 稽核人員建立正式註冊制度**，並規定嚴格的專業要求和獨立性標準。IAAA 在實際監管框架方面具有影響力。

03

推動研究與知識分享：

IAAA 舉辦研討會系列，邀請其會員分享研究成果。。

IAAA 的出版物中也涵蓋了對衡量大型語言模型（LLM）偏見的基準研究，以及如何透過 AI 審計讓青少年參與負責任 AI 實踐的探討。

04

引領業界討論與標準制定：

積極參與多個高知名度活動，爭取**推動演算法問責的討論**。

提出透過參與式稽核程序實現以社群為中心的 AI 問責方法。

積極參與國際合作，致力於**塑造人工智慧生命週期的全球標準**，以提高演算法透明度並推廣值得信賴的人工智慧實務。

05

倡導負責任的 AI 實踐：

IAAA 的使命是幫助**重新定義演算法稽核標準並確保負責任的人工智慧實務**。。

致力於為演算法的**透明度和公平性建立堅實的基礎**。。

IAAA 鼓勵其成員和廣大社群共同參與，**推動變革和創新，以實現負責任的 AI 發展**。

資料來源：2025 FEB IAAA-AI Auditing Update

© 2025 CAA All rights reserved.

ISACA AI 稽核工具概述

- 基於人工智慧(AI)的發展正以驚人的速度持續進行，當AI技術與業務需求的結合度越高，且也成為企業的產品和服務交付項目時，稽核程序與評估技術也必須是多面向且可動態調整，以跟上AI進步的幅度。因此，期待所有稽核人員遵循一套定制規則是不切實際的。

- ISACA於近期提出的AI稽核工具，是源自法律框架的AI控制庫。讓稽核人員身處在AI生命週期的不同面向，可以更容易理解且掌握相關控制風險；也讓電腦稽核師對於AI這門新興技術領域，可以有效地進行控制。

© 2025 CAA All rights reserved.

ISACA AI 稽核工具概述

- 提供一種結構化的AI合規性和評估控制的方法，可以讓AI開發人員與稽核人員有所依循。
- 企業可使用現有的AI控制來發展強大的AI評估程式。這些控制項目會被綜合歸類至不同的對照組，以便確認評估範圍。
- 控制系列和類別會跨越不同領域，包括對抗性防禦和穩健性、AI資料隱私和權利、AI智慧模型治理，以及AI道德治理與問責等。

© 2025 CAA All rights reserved.

AI 稽核工具

控制措施已反映到稽核人員可能已瞭解的法規和標準的現行指南。

政策一致性：評估控制措施是否符合相關監管要求、產業標準、地理區域和組織政策。

證據收集：定義證實合規性所需的證據類型，包括文件、日誌、配置設定和其他工作。

測試程序：指定在評估控制措施在減輕相關風險方面的有效性測試程序，包括模擬、評估和技術等方法。

文件和指導：提出詳細文件和指導資料以協助稽核人員和其他評估人員有效執行人工智慧控制評估。

© 2025 CAA All rights reserved.

AI 稽核工具

有效幫助使用者導航並利用此工具來確定AI控制的優先順序和評估。

需注意的是，稽核人員需自行決定評估現有的詳細流程，並確定某些控制是自動或手動，測試步驟是否適用於正在進行稽核的環境，或必須執行額外的測試。

可幫助稽核人員正在尋求支援AI評估工作的綜合資源，包括AI術語或客製化自我評估或稽核計畫、以解決演算法偏差、資料隱私違規等問題。

© 2025 CAA All rights reserved.

參考相對完善的安全框架與標準

- 每個AI控制的設計都是為了確保滿足可解釋性維度的AI目標：
  - 基本原理、責任、數據、公平、安全和效能以及影響。
- 控制措施與完善的安全框架和標準
  - 包括NIST 800-53 和ISO 27001相結合
  - 透過對選定的AI特定指南（包括歐盟AI法案）進行解構和分析
  - 可進一步瞭解資料，包括
    - 新加坡的AI治理框架模型、
    - MITRE (美國國防技術研究NGO)的 Atlas (對抗性機器學習威脅矩陣)
    - 和 OWASP 機器學習十大安全風險。
  - 依以上三類區分後，在稽核工具內的分類產出示意圖如下表。

© 2025 CAA All rights reserved.

業務影響與風險

組織需要確定如何遵守內部政策、法律和法規，以及解決AI採用所帶來的相互關聯風險所需的適當合規程度。

對於組織來說，制定可靠的計畫至關重要，以確保主動監控複雜且快速發展的法律和監管環境。

反映到企業當前流程和AI控制要求，應可以識別合規差距。

向利害關係人進行AI透明性報告，並對評估驅動的控制補救活動進行優先排序。

© 2025 CAA All rights reserved.

最低稽核技能

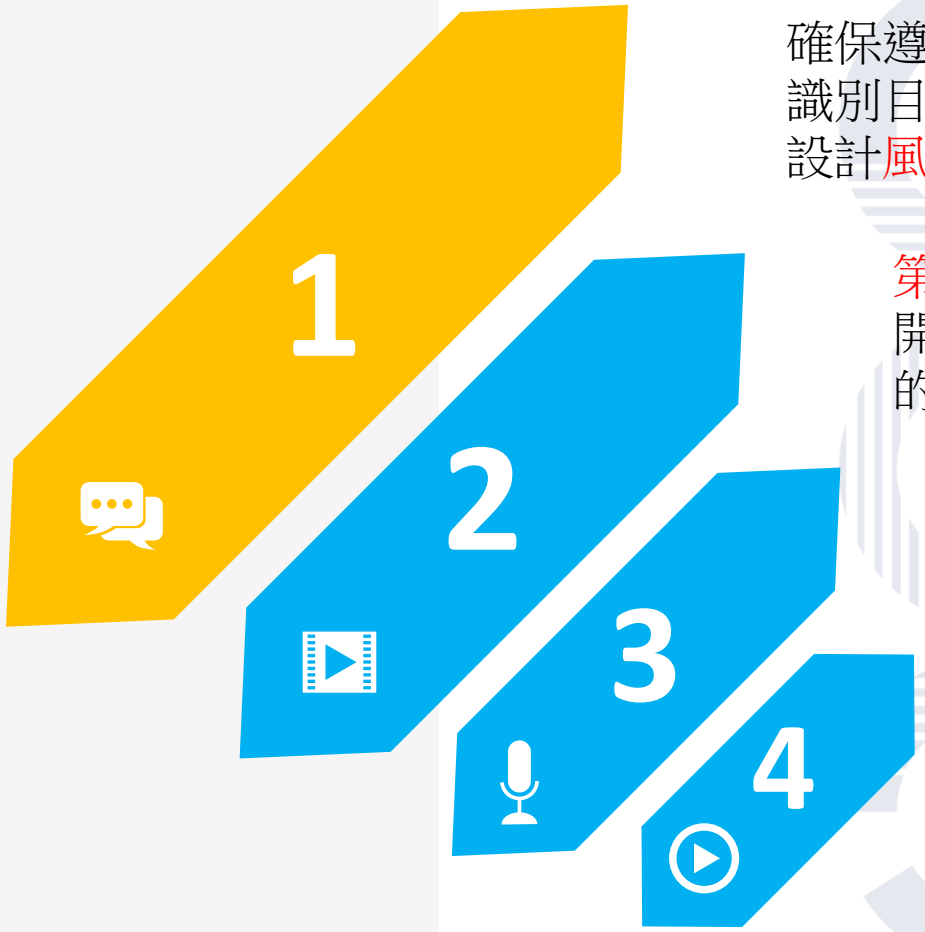
稽核專業人員必須瞭解安全、內部控制與技術流程，並能夠對那種評估方法或測試步驟適用於評估環境或是否執行額外測試程序進行專業判斷。

有效的AI稽核需要IT、數據科學、道德和行業或特定領域的專業知識，對於某些稽核員來說會是重大障礙。

稽核員應具備足夠的職能與業務知識，以確定與業務策略的一致性。執行此稽核的人員應驗證他們是否已進行所需的研究以理解AI技術性質及其相關風險。

© 2025 CAA All rights reserved.

利用AI框架強化三道防線



確保遵守AI法規需要公司對當前環境進行改變。首先是建立AI系統清單，識別目前使用的AI系統（包括由第三方管理且其輸出被公司使用的系統），設計風險分類流程並將其與當前流程整合。

第一道防線可以使用建議的證據/可交付成果/工件來增強目前流程或開發新流程。同時應創建一個識別、管理和處理AI系統所產生的風險的流程。

第二道防線應確保第一道防線制定政策、程序和控制措施並對其充分性進行自我評估。可以更好地了解如何設計政策/程序/指南來實現預期目標。

第三道防線負責了解AI系統產生的風險，並將其作為執行基於風險的內部稽核計畫的一部分，以測試相關控制的完整性和運作有效性。可以使用建議的評估方法來了解如何評估和測試證據、可交付成果、工件以及控制的運作有效性。

© 2025 CAA All rights reserved.

Next Step

每一項變革的採用都取決於多種因素，其中最重要的是人才。

在等待新法規生效的同時，企業應該重新思考和設計AI人才管理，以確保使用AI系統的員工擁有適當的知識。

ISACA 最近的AI研究表明，40% 的組織根本不提供AI培訓，大多數受訪者認為自己的技能方面有人才缺口。

未來的監管可能會改變，但有一件事是肯定的：AI將留在我們身邊。

依據AI相關合規與監管法令的不斷更新，持續優化適用於自己組織的AI稽核工具。

© 2025 CAA All rights reserved.

AI證照-以ISACA為例

ISACA規劃的AI證照有三種:

- 一是為稽核人員設計的 Advanced in AI Audit (AAIA),
- 二是為資安管理人員設計的 Advanced in AI Security Management (AAISM),
- 三是為風險管理人員設計的 Advanced in AI Risk, 皆已開放報考;
- 相關說明連結如下：
  - https://www.isaca.org/credentialing/aaia
  - https://www.isaca.org/credentialing/aaism
  - https://www.isaca.org/credentialing/aair
  - **ISACA Artificial Intelligence Audit Toolkit** https://store.isaca.org/s/store#/store/browse/detail/a2S4w000007kB9pEAE

© 2025 CAA All rights reserved.

ISACA的其他AI治理白皮書

ISACA白皮書下載專區

https://www.isaca.org/resources/insights-and-expertise/white-papers

- 人工智慧革命的機會與挑戰：風險管理

https://www.isaca.org/resources/white-papers/2026/the-promise-and-peril-of-the-ai-revolution

- 利用 COBIT 實現有效的 AI 系統治理

https://www.isaca.org/resources/white-papers/2025/leveraging-cobit-for-effective-ai-system-governance

- 瞭解歐盟人工智慧法案：要求和後續步驟
- https://www.isaca.org/resources/white-papers/2024/understanding-the-eu-ai-act

© 2025 CAA All rights reserved.

THANK YOU

© 2025 CAA All rights reserved.