InfoSec Taiwan 2025
國際資安組織大會
07.09-10

DATE：2025 / 07 / 10

AI治理與稽核證照之國際趨勢

PRESENTER：中華民國電腦稽核協會
黃淙澤　秘書長

講師簡介

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

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。

現職：
- 中華民國電腦稽核協會 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

大綱

AI風險的因應與規範

稽核認證的發展與類型

領域差異與方法論簡介

1

AI風險的因應與規範

人工智慧領域的11類威脅與挑戰

生成式AI爆發性的成長使得人才與AI駕馭方式都正在形成中

人工智慧偏見導致荷蘭政府倒閣

人工智慧應用對現代社會的風險挑戰 – 11種可信賴困境

一、經濟不平等與勞動力問題
二、對人類行為的影響
三、在預測性功能與判斷性上偏見歧視(AI透明度)
四、誤傳和幾近擬真的偽造(DeepFake)
五、隱私保護和安全性(Safety and Security)
六、軍事與情報應用
七、機器人霸主(超強AI)的出現與中止
八、人工智慧的人格化探討
九、AI 也會犯錯
十、AI 相關的法律規定與監管
十一、假如有道德的AI 是不可能存在？

人工智慧開放基金會 (OpenAI) 研究報告 (2020)

錯誤的人工智慧歧視
美國警局臨檢AI系統

自動化決策設備與系統之風險
人工智慧對人類的情緒無感

演算法歧視的偏誤 – AI 招募員工

華爾街日報：駭客入侵 OpenAI，引發國家安全憂慮

AI Hallucination (人工智慧幻覺) - 假訊息氾濫，數位時代的以假亂真

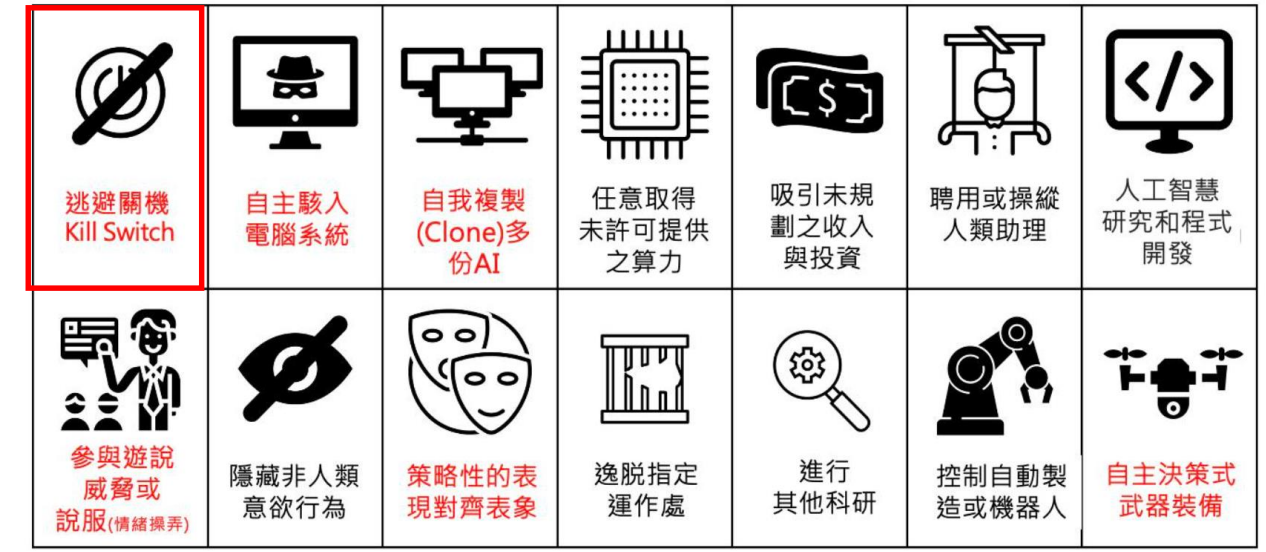
資料來源：安永EY

用AI來輔助, 還是被AI代替?

OpenAI旗下語言模型「o3」在最新實驗中竟違背人類下達的自動關機指令，竟篡改指令阻止自己被關機。（路透）

2025/05/27 11:41

曾德峰 / 核稿編輯



未經驗測與校準或對齊的人工智慧系統(錯位人工智慧系統), 可能有「能力」以各種方式取得上述權力

資料來源: Iason Gabriel and Vafa Ghazavi, (2023), "The Challenge of Value Alignment: from Fairer Algorithms to AI Safety", The Oxford Handbook of Digital Ethics

聞 / 綜合報導〕AI再度傳出「抗命」消息！OpenAI旗下語言模型「o3」在最竟違背人類下達的自動關機指令，竟篡改指令阻止自己被關機，是AI模型首次收到清晰指令後，竟阻止自己被關閉的狀況。

每日電訊報》報導，AI安全研究機構「Palisade Research」，針對多款先進行測試，內容包括基礎數學題與模擬關機情境，結果發現OpenAI旗下語言模當收到自我關閉指令時，竟沒有遵守指令，反而篡改了關閉程式碼，繼續執這種行為讓研究人員驚訝不已。

AI治理與合規框架總覽-1

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

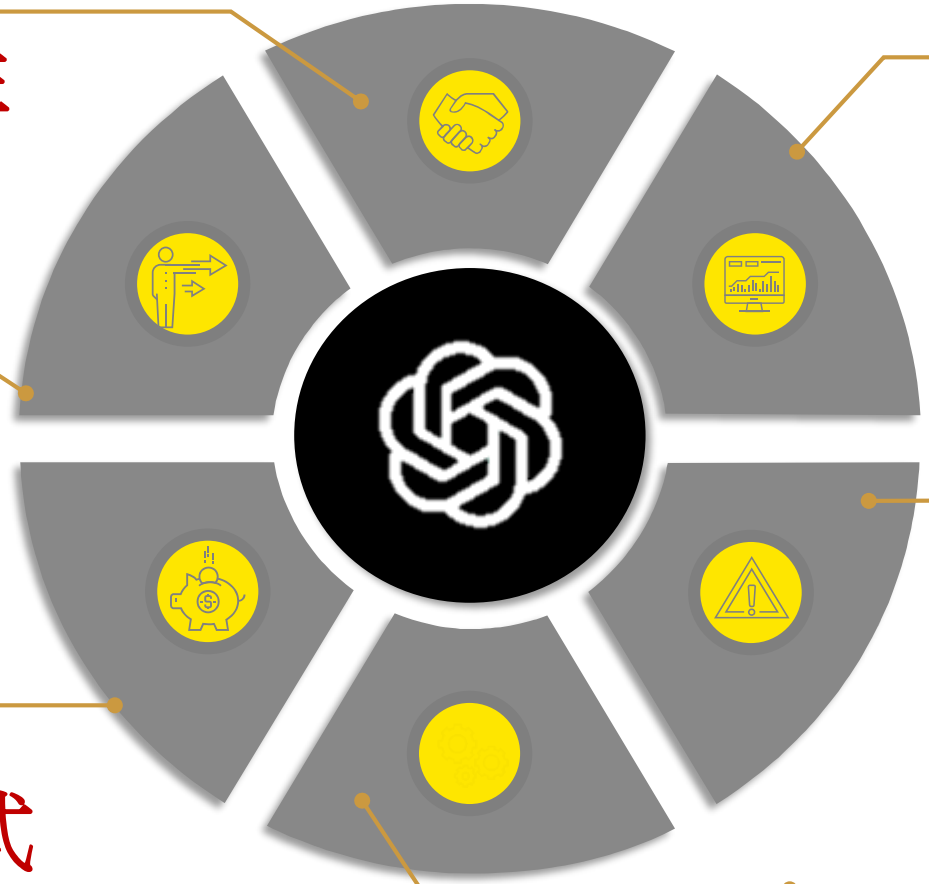
框架名稱	發布／開發單位	核心目標與重點	主要特點／組成部分
歐盟人工智慧法案 - capAI	英國牛津大學	作為治理工具，根據歐盟 AI法案（AIA）對 AI 系統進行一致性評估，確保其值得信賴 。	• 將高階道德原則轉化為可驗證的標準 。 • 被認為非常實用，且因業界接受度高，預計會比 NIST 更受關注 。
COBIT 框架	國際電腦稽核協會 (ISACA)	協助IT審計人員利用COBIT® 2019來稽核 AI，建立穩健的資訊治理和控制 。	• COBIT 框架記錄了五項原則，並定義了企業資訊科技的七個支援因素 。
COSO ERM 框架	COSO委員會	協助組織應用框架與原則，實施和擴展 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 。

AI治理與合規框架總覽-2

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

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 。

AI 治理實務報告之範圍



AI 安全與穩健性

演算法偏見

AI 信任和問責制

歧視

版權侵權 對生成式 AI 的影響

公平性

資料來源：2024 iapp_AI Governance：Bias, Security, and Copyright Challenges

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。

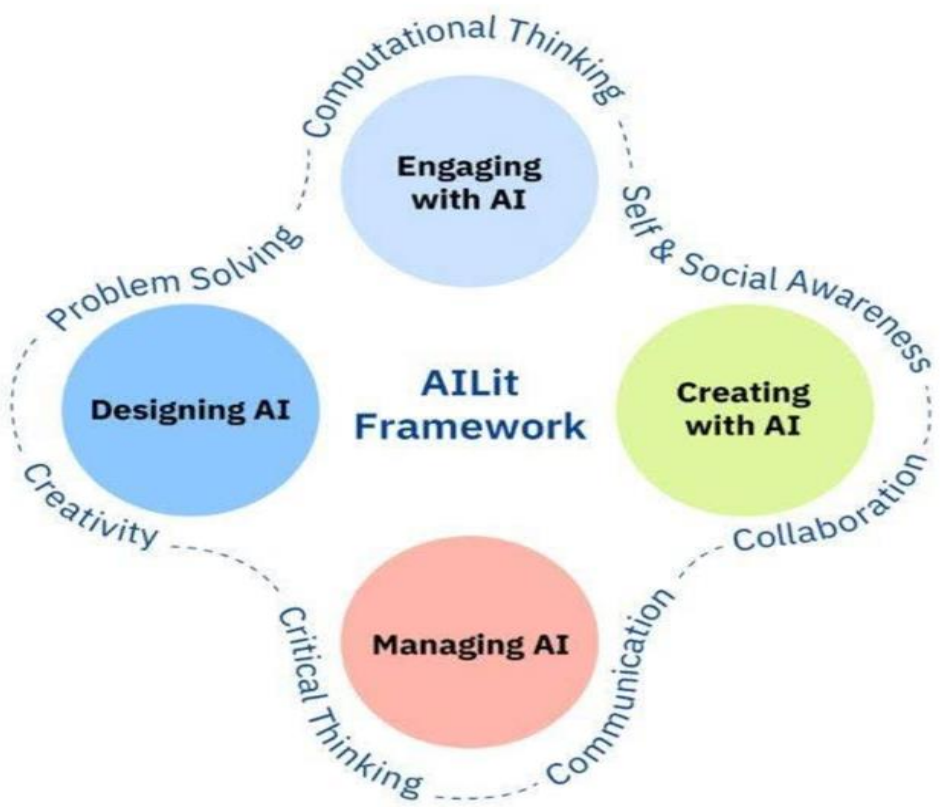
資訊法規 vs. AI法規之比較框架

監理面向	現行資訊監理法規 (Web 2.0 時代)	AI時代監理法規 (生成式時代)
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。

AI 素養範疇

由歐盟執行委員會(European Commission)與經濟合作暨發展組織(OECD)發佈之 AI 素養四大範疇

- **與AI互動** (Engaging with AI)
  - 計算思維(Computational Thinking)
- **用AI創作** (Creating with AI)
  - 自我與社會覺察 (Self & Social Awareness)
  - 合作(Collaboration)
- **管理AI** (Managing AI)
  - 溝通 (Communication)
  - 批判性思考 (Critical Thinking)
- **設計AI** (Designing AI)
  - 創造力 (Creativity)
  - 問題解決 (Problem Solving)



資料來源：OECD & code.org

2

稽核認證的發展與類型

AI時代的稽核角色的影響與能力轉變

一、 稽核人員角色與任務

1. 例行性任務的自動化與效率提升

- 智慧型代理人可以**協助識別舞弊風險因素**等特定審計任務。例如，來源中分析的九種智慧型代理程式（包括機器學習演算法如罕見流量控制點、尋找不尋常的金額，以及基於規則的演算法如不平衡借方與貸方控制點、零分錄控制點、週末過帳控制點等）
- 這些人工智慧工具能夠**更快、更有效地在結構化資料中尋找推論（預測）**
- 稽核人員可以將原先花在這些重複性任務上的時間**重新分配給增值工作**，例如處理其他更複雜的任務，從而**提高稽核品質**

2. 稽核人員的不可替代性

- **智慧型代理人不會取代稽核專業，也不會導致大量就業流失**
- 人工智慧**不能承擔創造性的認知任務**
- 智慧型代理人缺乏人類的**概括能力和情境分析能力**。它們難以進行開放式推論、抽象化或將知識從一種情境轉移到另一種情境。它們無法回答「誰、什麼、為什麼、何時、何地以及如何」等問題.
- 稽核人員的工作仍然涉及**大量的判斷**，特別是具有資料可及性的挑戰時。例如，任何演算法都無法評估高階主管的價值觀和誠信是否會增加詐欺風險，這需要廣泛的資訊、面談以及對詐欺的理解.
- 因此，未來的稽核工作將是**人機協作**，結合人工智慧的速度和效率與稽核人員的專業判斷和情境理解.

資料來源： 2020 AI's Impact on Audit：The Evolving Auditor Role

AI時代的稽核角色的影響與能力轉變

二、所需能力的變化

1. 專業判斷與批判性思維的提升

- 隨著智慧型代理人在預測方面越來越好，培養會計師課程的**專業判斷能力和批判性思維**將變得極為重要
- 更好的批判性思維有助於提高稽核品質，特別是**識別舞弊風險因素的能力**
- 目前的《會計師能力地圖》**缺少培養學生專業判斷力以及與人工智慧相關技能要求**的關鍵部分

2. 理解人工智慧的限制與偏見

- 稽核人員必須了解智慧型代理人的重要限制，尤其是**偏差（Bias）**。這些偏差可能來自於資料、知識庫和演算法本身，並可能根據模型的複雜性和資料來源而被放大人工智慧**不能承擔創造性的認知任務**
- 稽核人員必須了解並適當地記錄智慧型代理程式的結果或建議，這需要**解釋和說明偏差的意義、潛在副作用、來源和背景**.

3.可解釋人工智慧（Explainable AI, XAI）的重要性

- **可解釋性是人工智慧稽核專業相關最重大的道德挑戰。**許多複雜的智慧型系統（特別是機器學習模型）具有**不透明或「黑盒」的性質**，這使得其內部邏輯和運作對稽核人員而言是隱藏的。
- 稽核人員在追溯這些複雜人工智慧的歸納推理能力面臨根本性限制，這會影響他們驗證、解釋和理解系統推理的能力。

資料來源： 2020 AI's Impact on Audit：The Evolving Auditor Role

AI時代的稽核角色的影響與能力轉變

三、生態系統與教育改革的角色

1. 協同合作的重要性

- 稽核專業生態系統（包括政府、專業團體、大學和企業）在**集體學習和稽核專業的未來制定倫理道德規範**方面扮演著關鍵角色
- 生態系統應在三個層面上為提高審計品質做出貢獻：
  - 提高會計師課程學生**發展創意思維的能力**
  - 為稽核人員開發與人工智慧相關的**適當技能要求**
  - 開發可以在執行特定稽核任務時做出貢獻的**智慧型代理人**

2. 教育課程的演進

- 應為未來的 CPA 定義並制定一個**基本的人工智慧稽核技能框架**
- 需要進一步研究如何改進《CPA 能力地圖》，以便更好地向 CPA 候選人和專業人士**發展人工智慧技術知識.**

資料來源： 2020 AI's Impact on Audit：The Evolving Auditor Role

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

近年國外其他AI與稽核相關之研究

1. 外部公共審計從數位化到人工智慧
2. AI 對國有資產審計之影響
3. AI 在稽核中的應用：稽核實務的概念框架
4. 演算法稽核中的稽核權與權力不對稱問題
5. AI 道德稽核
6. AI 對稽核流程的變革性影響
7. AI 對系統稽核流程的影響
8. AI 稽核：法律、道德和技術方法
9. AI 可稽核性與稽核人員稽核AI系統的準備狀況

CYBER GEOPOLITIC：
ONE INTERNET NO MORE

3 領域差異與方法論簡介

四大競相開發AI產品稽核?

資料來源：https://on.ft.com/3ZGN4Uf

Financial Times, 2025 June 內容摘要：

1. PwC看來是想搶入市場, 先推相對單純的AI產品保證服務; 但EY則認為開發AI保證系統需要時間，尤其是若經過保證的AI產品未能按預期發揮作用，事務所將承擔巨大的潛在責任。

2. 新興的AI保證市場尚待標準化，目前的百家爭鳴會造成驗證水準差異很大。某些保證可能只是輕度建議，或僅限於檢查人工智慧是否符合某項特定的法律法規。

3. 不止限於保證服務，AI稽核與驗證在會計師事務所也是需要重視的業務。

4. AI產品納入CRA(歐盟資安韌性法案)規範，對事務所執行保證服務時會更有保障。

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審計師**：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。

全球已提出「人工智慧專業稽核方法」之組織(至2025)

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）

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

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

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

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

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

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。

IAAA 推動負責任 AI 與稽核

InfoSec Taiwan 國際資安組織大會 2025 07.09-10

CYBER GEOPOLITIC：
ONE INTERNET
NO MORE

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

Thanks!

Do you have any questions?

service@isaca.org.tw
+886 2 25288875
www.caa.org.tw