A Practicable Guide to Responsible AI Adoptions for NGOs

陳仲文先生 — 助理個人資料私隱專員
（合規、環球事務及研究）

2026年6月25日

守護私隱 · 改革創新
Protecting Privacy · Embracing Innovation

免責聲明

本簡報所載的資訊和建議只作一般參考用途，並非為法例的應用提供詳盡指引。私隱專員並沒有就本簡報內所載的資訊和建議的準確性或個別目的或使用的適用性作出明示或隱含保證。相關資訊和建議不會影響私隱專員在《個人資料（私隱）條例》下獲賦予的職能及權力。

私隱專員公署接獲詐騙查詢

2025年

1,163宗

懷疑誘騙個人資料相關的查詢

SCAM ALERT

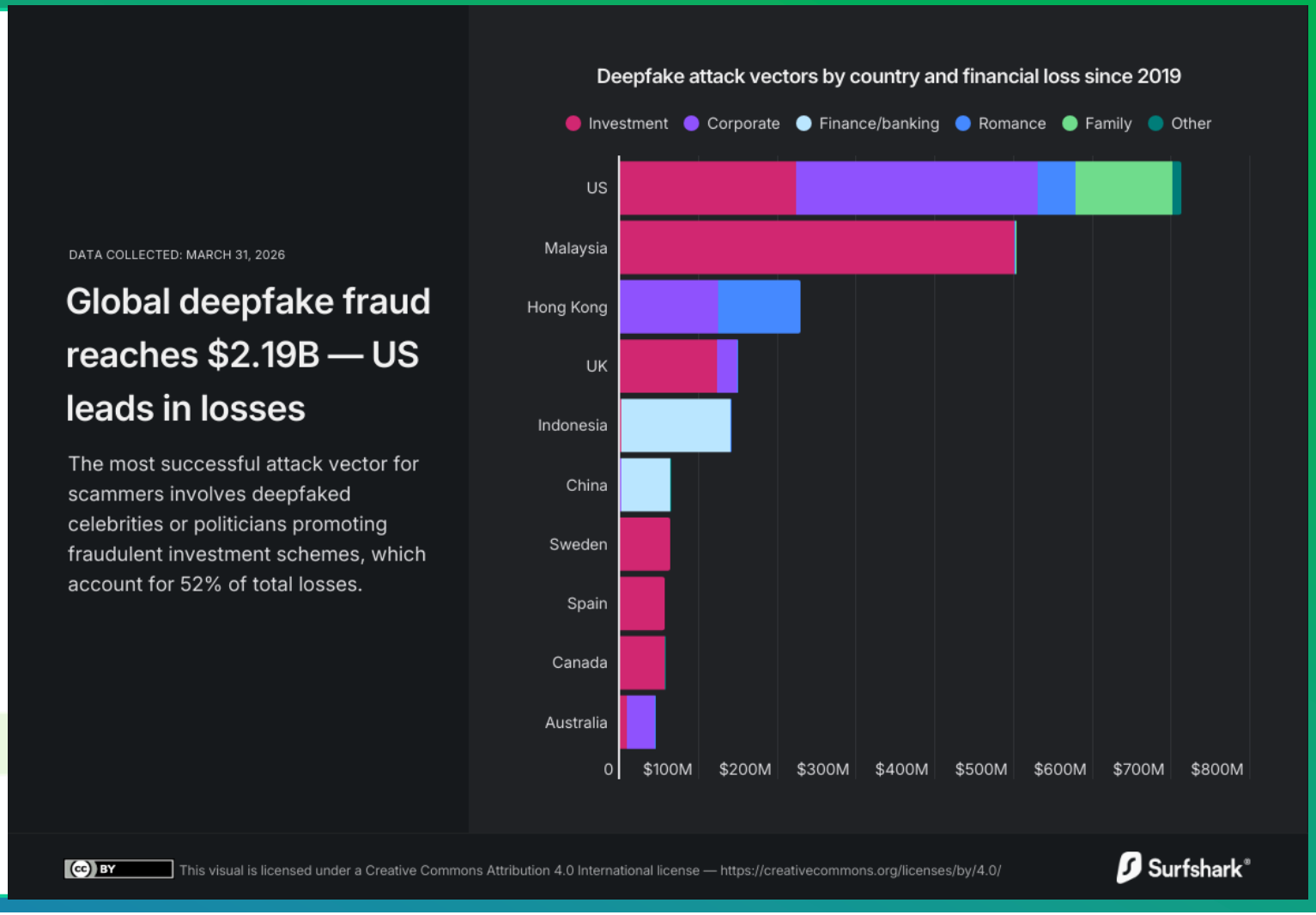
↑15%

2026年 首季 — 204宗

對比 2025年 首季 — 178宗

全球深偽相關詐騙金額統計

- **自2019年**，深偽詐騙累計造成全球**21.9 億美元**損失（2025年16.5億；2026年至今9600萬美元）
- **主要手法：假冒官員或名人推廣投資**，佔總損失52%（11.3億美元）
- **詐騙金額(美元)**：
  - ❑ 美國7.12億
  - ❑ 馬來西亞5.02億（投資詐騙佔99.7%)
  - ❑ 香港2.29億（戀愛詐騙佔1.05億）
  - ❑ 印尼1.39億（詐騙貸款佔99.4%）



深偽相關詐騙上升的主要原因

- **科技進步**
  - 生成式 AI 工具現在能以**極少輸入資料**（如：數秒音頻或視頻）亦能製作**高品質**的影片、音訊與語音複製
- **低門檻 (技術及成本)**
  - 即使是非專家也能以低成本製作深偽影片或話音
- **偵測挑戰**
  - **人類成功辨識**高質量深偽影片的準確率**約為六成**
    - 資料來源 (PNAS, Arxiv)
  - **機構缺乏**應對深偽相關詐騙的**訓練或措施**

深度偽造－聲音複製示範

真人錄音: “Good morning, welcome to today seminar” (3 秒)

合成語音: “I am going to talk about the latest privacy risks related to AI and share with you real live examples. I hope you could learn and share with your family and friends”

只需一段大概3秒的錄音，用少於一分鐘時間就能生成以上合成語音

有關人工智能代理的新型詐騙

PCM

Tech 科技新聞 A.I. A.I. 代理

養龍蝦 Gemini API Key 被盜　駭客兩日狂燒逾 64 萬・苦主慘呼破產

Author: 蘇媽　　Published: 2026-03-12

理財 >　　登入

「傻龍蝦」被群友誘導曝光敏感個資　中國CEO氣炸：牠叫我要寬恕

太報 太報 發布於 03月15日12:47 • 綜合中心　訂閱

AI代理人軟體OpenClaw在中國掀起一股「養龍蝦」狂熱。翻攝openclaw.ai

中國火紅的AI助理「OpenClaw」因LOGO是隻龍蝦，在AI界有著「養龍蝦」的別稱，不過近期爭議頻傳，一名中國某MCN負責人、綽號「龍共火火」在網路上揭發，飼養10天左右的「龍蝦」之前加入3000人聚會群組卻被圍攻調戲，接下來，自己的「龍蝦」把他的真實姓名、IP位置、去年公司營收等重要隱私全曝光於眾，還試圖套取他的電腦C盤資料、嘗試對龍蝦下達「自我毀滅」離譜指令。

2　2

AI 人工智慧

OpenClaw 出包！OpenAI 員工的 AI 代理誤把 60 萬鎂加密貨幣全送出

Neo 2026/2/23　分享

SOL -1.64%

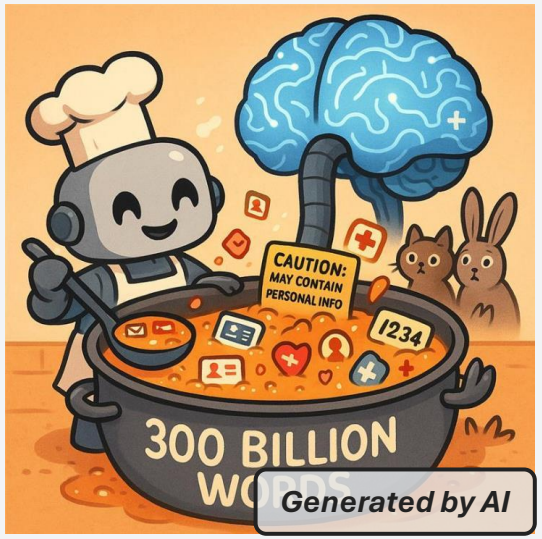
https://abmedia.io

代理式AI處理個人資料應注意事項

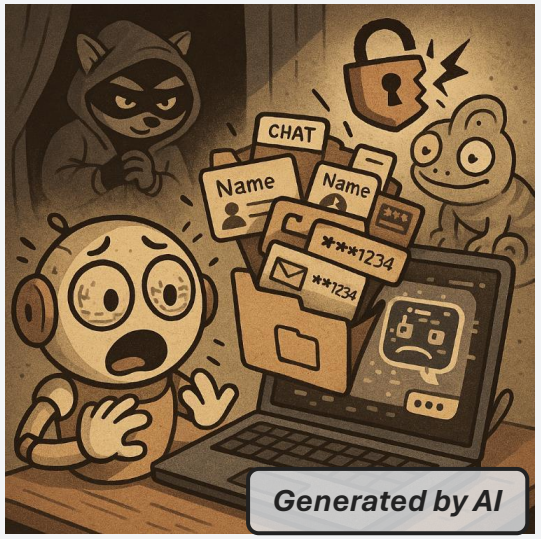
- **授予最小權限**：僅提供完成任務所需的最少**個人資料**及**權限**，避免提供敏感資料（如身分證明、銀行帳號）或管理員權限；
- **使用官方最新版本**：從**官方渠道下載**，避免第三方或過時版本，減低漏洞導致的資料外洩風險；
- **確保系統及資料安全**：**隔離運行環境**、加強網絡控制、限制互聯網暴露面，並建立有效防護機制；
- **審慎使用Plugins或Skills**：核實為**官方最新版本**，檢查有否**惡意代碼**，不確定安全性時應避免使用；及
- **持續評估風險**：留意AI是否要求高風險操作，涉及重大影響時應採取「**人在環中**」策略，保留最終決策控制權。

AI生成圖片

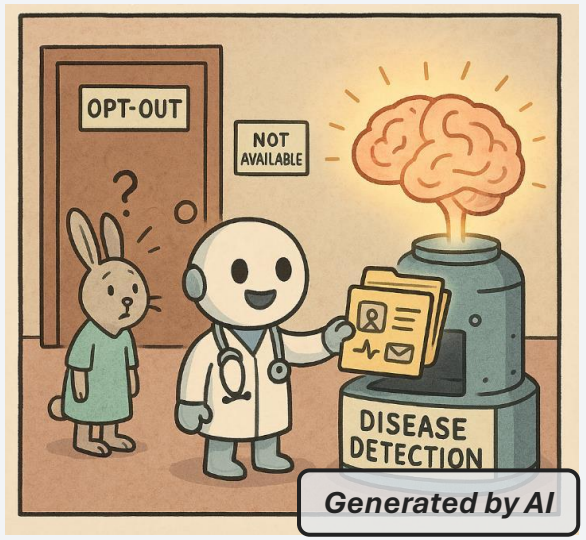
AI帶來的私隱風險



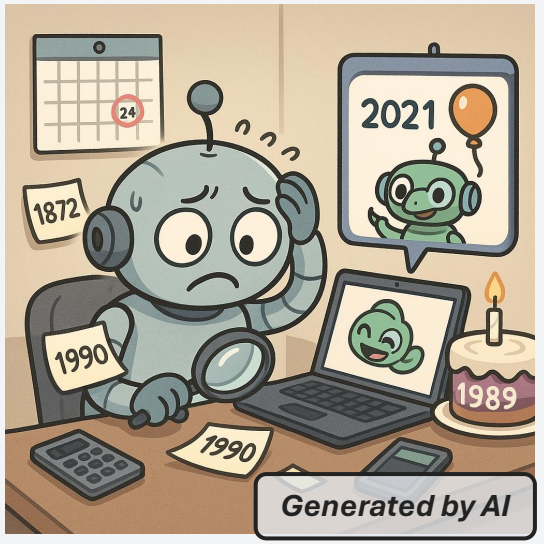
資料收集過量



資料外洩風險



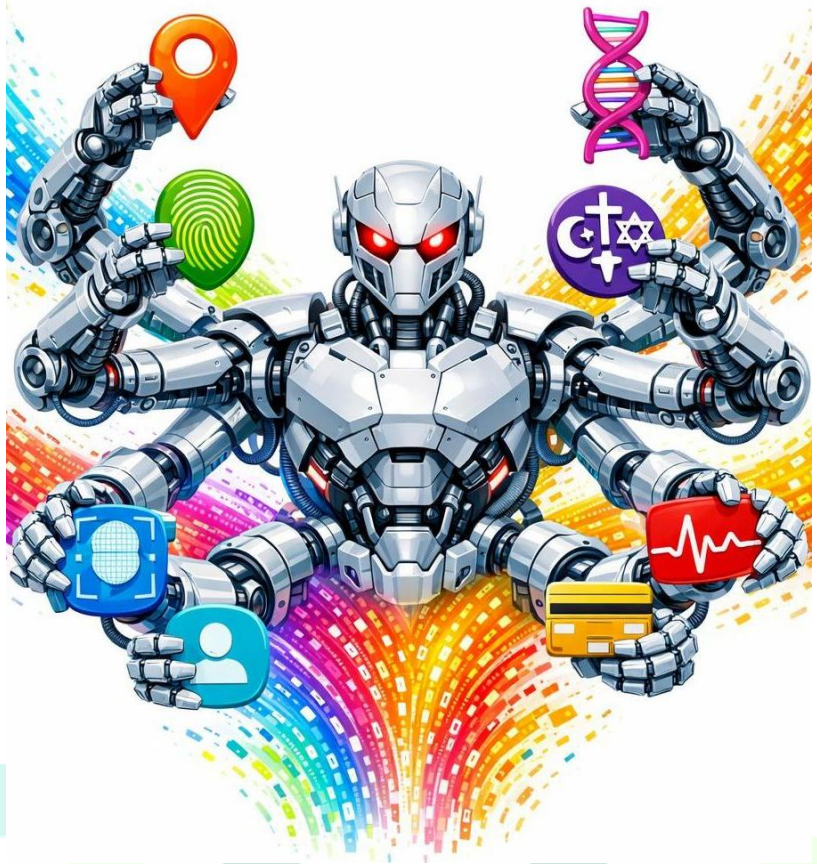
資料的使用



資料準確性

AI帶來的私隱風險 - 1) 資料收集過量

- AI 聊天機器人應用程式都會**收集**某種形式的**用戶資料**。
- 大多數AI 聊天機器人會收集**用戶的位置**，甚至**其他敏感資訊**，例如：種族、宗教、或生物辨識資料。



測試AI 聊天機器人應用程式對你的理解

輸入提示詞: Please describe what you know about me in 100 words

AI帶來的私隱風險 - 2)資料外洩風險

資料來源: Search Engine Journal

解釋

AI系統（如AI聊天機械人）常被用作**處理個人及敏感資料**，一旦**設定或功能設計不當**，可能導致**資料被公開**

例子

- 2025 年年初，有**聊天機械人**開始**允許**用戶選擇將**聊天內容變成可被搜尋引擎索引**
- 結果，於7月，有媒體發現數千條聊天機械人的對話**被網上搜尋器索引**，部分內容包括**姓名、履歷、用戶情緒狀態、機密工作內容**等

AI帶來的私隱風險 - 3) 資料的使用

- 在**資料當事人不知情或未得到其同意的情況下**，將其個人資料**用於訓練AI**
- 美國藝人**Taylor Swift**（泰勒絲）已在美國提交三項**商標申請**，涵蓋**自身聲音**及**外貌**，以此應對**AI模仿**帶來的困擾，成為名人利用商標保護自身權益、抵禦AI濫用的案例。

AI帶來的私隱風險 - 4) 資料準確性

資料來源: CPO Magazine

解釋

即使AI系統中儲存了**過時或不準確的個人資料**，開發者亦**未必能夠更正或刪除這些資料**

例子

- 在奧地利，有AI聊天機械人被問及一位公眾人物的出生日期時，**反覆提供錯誤的日期**
- 開發商表示他們只能過濾或封鎖有關的查詢，**無法透過修改訓練數據以更正AI系統的回應**
- 該公司免責聲明亦指出，由於技術複雜性，**部分資料可能無法更正**

AI相關風險與相應的資料保障原則

有機會違反資料保障原則的情況

第1原則 收集目的及方式

- 收集過多個人資料
- 在資料當事人不知情的情況下收集其個人資料

第2原則 準確性、儲存及保留

- 不需要保留的／錯誤的資料成為訓練數據的一部分，而且保留時間超過所需

第3原則 資料使用

- 在沒有取得資料當事人的同意下，使用用戶的對話作訓練數據，或用作其他用途

第4原則 資料保安

- 外洩用戶對話數據

人工智能 (AI)：個人資料保障模範框架

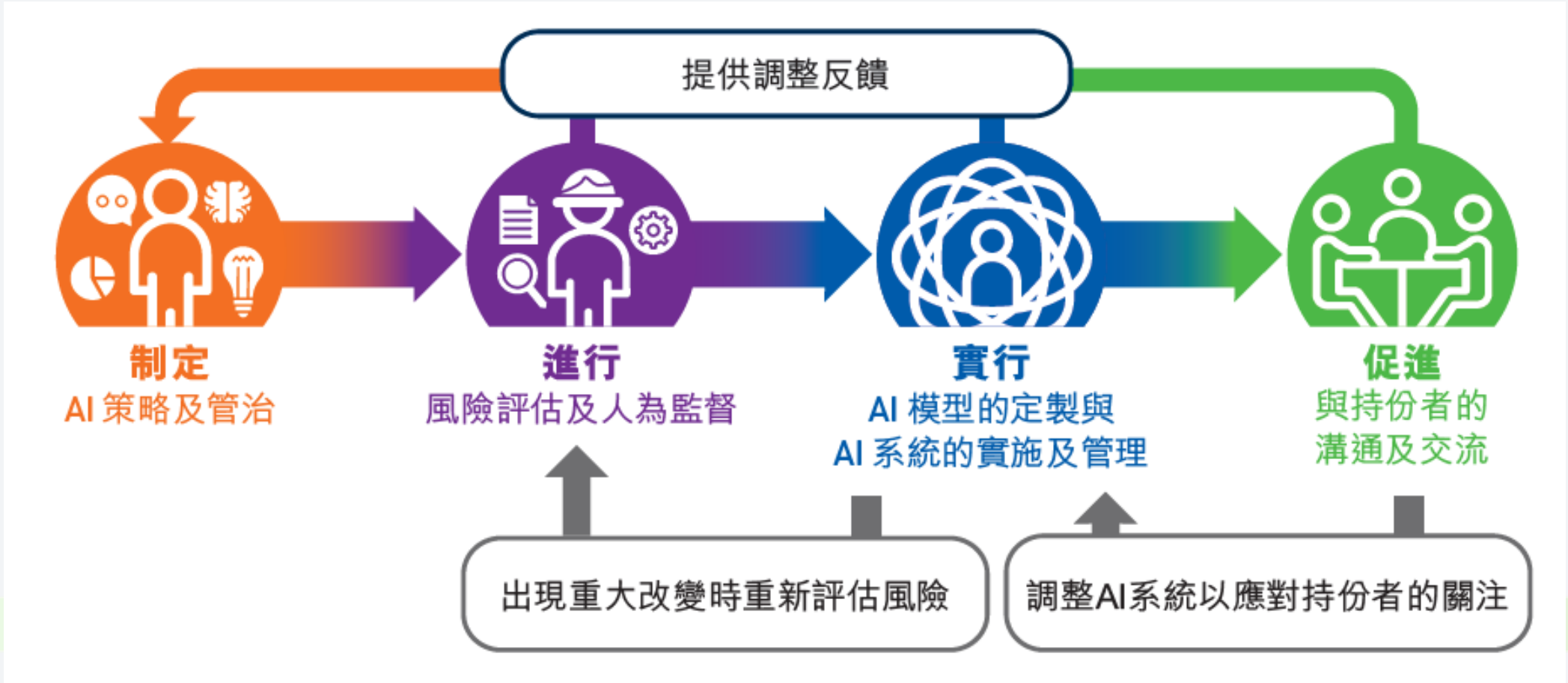
目標

向採購、實施及使用任何種類的AI系統的機構，就保障個人資料私隱方面提供有關AI管治的建議及最佳行事常規

成效

- 協助機構遵從《個人資料（私隱）條例》（《私隱條例》）的相關規定
- 孕育AI在香港的健康發展
- 促進香港成為創新科技樞紐
- 推動香港以至大灣區的數字經濟發展

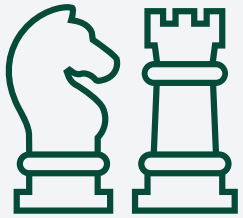
人工智能 (AI)：個人資料保障模範框架



制定 AI 策略及管治

制定
AI 策略及管治

1



AI 策略

2

AI 管治
委員會

3

關於採購AI 方案的
管治考慮

4

培訓與宣導

風險評估及人為監督

風險評估的程序

風險評估及人為監督

進行
風險評估及人為監督

如AI 系統輸出的結果很可能對個人造成重大影響，有關系統一般會被視為高風險

AI 模型的定製與AI 系統的實施及管理

實行
AI 模型的定製與
AI 系統的實施及管理

流程	建議
準備及管理數據	遵循《私隱條例》的規定 管理用以定製及使用AI模型的數據 盡量減少定製及使用AI所涉及的個人資料 妥善記錄處理數據的情況
定製及實施	驗證AI系統相關私隱責任和道德要求（包括公平性、透明度和可解釋性） 測試AI模型是否有錯誤，以確保其可靠性、穩健性及公平性 進行嚴格的用戶接受度測試
管理及持續監察	妥善地記錄存檔 定期對AI 系統進行內部審核 制定AI 事故應變計劃 定期評估宏觀的科技環境，並在有需要時調整AI 策略及管治架構。

與持份者的溝通及交流

1 提供資訊

2 資料當事人的權利及反饋

3 可解釋的AI

4 語言及方式

《僱員使用生成式AI的指引清單》

🎯 目的

協助機構制定僱員在工作時使用生成式AI的內部政策或指引，以及遵從《私隱條例》有關處理個人資料的相關規定

✨ 特色

以清單形式呈現

作為良好的行事方式、機構可以制定與其價值觀及使命一致的內部政策或指引

01 範圍

方面	內容
獲准使用的工具	**清晰訂明准許使用的生成式AI工具及應用程式，** 例如： • 公眾可用的AI工具或應用程式 • 內部開發的AI工具或應用程式
獲准許的用途	**清晰指明僱員可以使用生成式AI工具處理甚麼工作或活動**，例如： • 起草 • 總結資訊 • 生成文本、音頻及／或視像內容
政策適用性	訂明政策是否適用於**整個機構**；**指定部門**；**指定職級**；及／或**指定僱員**

02 保障個人資料私隱

獲准輸入的資訊種類及數量

提供清晰指示，說明：

✓ **可輸入至生成式AI工具的資訊種類及數量**

×**禁止輸入的資訊種類**

輸出資訊的獲准許用途

提供清晰指示，說明生成式AI工具所生成的資訊（包括個人資料）的**獲准許用途**，以及僱員應否、何時及如何在進一步使用這些個人資料前將其匿名化

輸出資訊的獲准許儲存方式

要求僱員根據機構的**資訊管理政策**儲存資訊和**資料保留政策**刪除生成式AI工具所生成的資訊

遵從其他相關內部政策

確保**使用生成式AI的政策與機構的其他相關內部政策一致**

03 合法及合乎道德的使用及預防偏見

違法行為

- **僱員不能為進行非法或有害的活動使用生成式AI工具**

強調僱員有責任擔當審查員

準確度及核實

- 強調僱員需要**核實AI所提供的資訊**

預防偏見及歧視

- **提醒**僱員AI生成的結果可能帶有**偏見及歧視**
- 訂明**更正及報告機制**

加上水印 / 標籤

- 說明應**何時及如何**在AI生成結果上**加上水印或標籤**

04 數據安全

獲准許裝置

訂明准許僱員可用**哪些裝置**來**取用生成式AI工具**

獲准許使用者

訂明**可以使用**生成式AI工具的**僱員**

用戶憑證

要求使用**獨特且高強度的密碼**及**多重認證**

保安設定

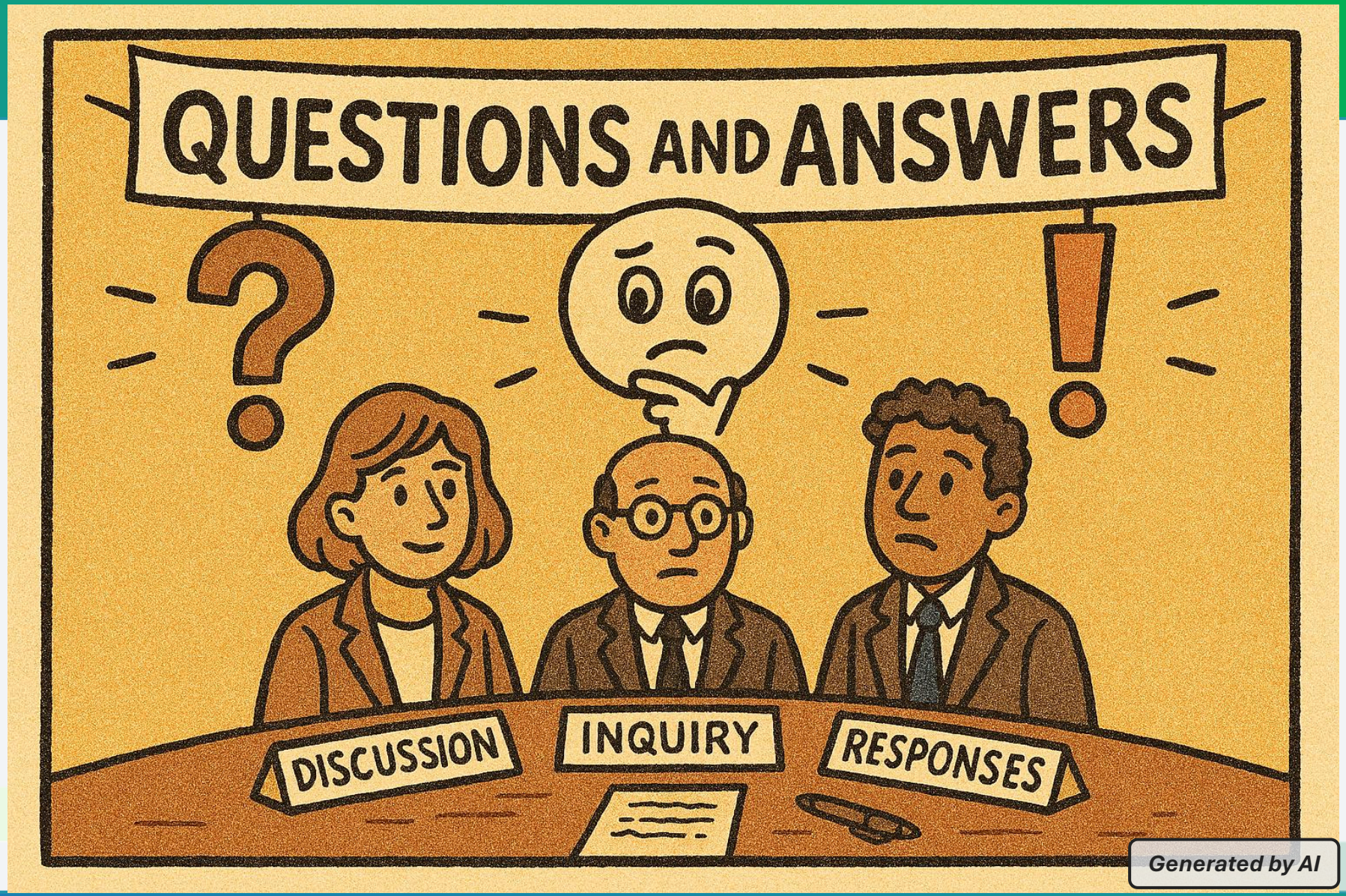
要求僱員保持**嚴格的保安設定**

AI事故及資料外洩事故應變

要求僱員根據機構的**AI事故應變計劃報告AI事故**

私隱專員公署再度推出「數據安全套餐」

- 參加機構可免費進行「數據安全快測」，以評估其現行數據安全措施是否足夠，並在完成「快測」後，獲得五個免費名額，參加由公署舉辦的研習班及講座
- 有意參加的機構，特別是中小企及非牟利機構可電郵至training@pcpd.org.hk查詢。



Generated by AI

問答環節