PCPD
HK

PCPD.org.hk

香港個人資料私隱專員公署
Office of the Privacy Commissioner
for Personal Data, Hong Kong

人工智能(AI):
個人資料保障模範框架

支持機構：

中華人民共和國香港特別行政區政府
政府資訊科技總監辦公室
Office of the Government Chief Information Officer
The Government of the Hong Kong Special Administrative Region
of the People's Republic of China

Hong Kong Applied Science and
Technology Research Institute
香港應用科技研究院

目錄

前言

人工智能（AI）的應用領域廣泛，商機無限，現時不少企業都採用 AI 技術，從而開源節流、提升生產力。不過，AI 帶來的風險亦不容忽視。舉例來說，若訓練 AI 的數據不足或質素參差，可能令 AI 系統作出錯誤或帶有歧視的決定；若訓練數據包含個人資料，系統或會在輸出過程中洩露這些資料。

創新應用帶來的新風險，難免會對監管帶來挑戰。隨着 AI 急速發展，世界各地紛紛推出不同法規，例如歐洲議會於 2024 年 3 月通過*《人工智能法案》*，按 AI 系統的風險程度實施相應規管要求；國家亦已於 2023 年 7 月發布*《生成式人工智能服務管理暫行辦法》*，以促進生成式 AI 技術的健康發展和規範其應用。

筆者樂見個人資料私隱專員公署推出*《人工智能：個人資料保障模範框架》*，積極為本港企業指引方向，協助企業利用科技帶來的便利之餘，進一步保障個人資料私隱。這實在有助提升企業的 AI 治理水平，確保 AI 獲妥善使用。

這框架以風險為本的原則，為有意採購、實施及使用 AI 系統的本地企業提供既實用又詳盡的建議。框架覆蓋業務流程的各個部分，企業無論是採購現有的 AI 方案或按本身需要定製 AI 方案，均可在指引中取得落地的建議。筆者鼓勵企業在採購及使用 AI 系統時，應先參考這框架，並實踐框架內的措施，從而以保障個人資料私隱，確保使用安全、合乎道德而且負責任的方式運用創新科技。

適逢國家正快速發展新質生產力，並開展了「人工智能＋」行動，以科技創新驅動產業發展，私隱專員公署推出這指引，正好可以協助企業善用 AI 技術，促進產業創新及升級轉型，幫助推進香港數字經濟發展、加速建設香港成為國際創科中心，積極融入國家發展大局。

黃錦輝教授

全國政協委員
立法會議員
香港中文大學工程學院副院長（外務）

2024 年 6 月

序

人工智能（AI）的突破性發展，正以超乎我們想像的方式徹底改變世界。我和個人資料私隱專員公署（私隱專員公署）的團隊堅信，雖然 AI 是一把雙刃劍，然而，在採取適當的保障措施下，例如實施全面的個人資料保障框架，AI 的使用將可以帶來更大的益處。作為保障個人資料私隱的守護者，我們致力倡導以合乎道德、負責任和私隱友善的方式使用 AI，令 AI 可持續健康的發展。

2021 年 8 月，私隱專員公署出版了*《開發及使用人工智能道德標準指引》*，這是亞太地區關於 AI 的具指導性的領先指引之一。我們明白 AI 為全球帶來的挑戰需要尋找一個全球性的解決方案，因此我們致力在國際層面參與討論，包括主辦 AI 國際會議，促進專家之間具意義的對話，並在環球私隱議會（一個匯聚超過 130 個資料保障機構的論壇）上與其他成員一同提出有關負責任及值得信賴的 AI 的決議。最近，聯合國大會通過歷史性的決議，推動「安全、可靠和值得信賴」的 AI，而國家早前亦發布*《全球人工智能治理倡議》*，可說為全球就 AI 的使用制定全面的個人資料保障框架積聚動力！

為體現國家的*《全球人工智能治理倡議》*，私隱專員公署制定了*《人工智能：個人資料保障模範框架》*（《模範框架》），以協助機構在採購、實施及使用 AI 系統時處理個人資料。這《模範框架》吻合一般業務流程，以確保 AI 系統的有效管治，並遵守私隱專員公署在 2021 年 AI 指引中倡導的三項數據管理價值和七項道德原則。《模範框架》提供了國際認可、切實可行及逐步式的建議，協助機構在利用 AI 的優勢之餘，亦保障個人資料私隱。

《模範框架》得以制定，有賴政府資訊科技總監辦公室和香港應用科技研究院的鼎力支持。我非常感謝各持份者，包括私隱專員公署科技發展常務委員會的成員和業界專家，向我們提供了寶貴的意見和觀點。我亦衷心感謝我的團隊，特別是蕭頴思女士、廖雅欣女士、陳筠朗女士和張偉瑜先生在草擬過程中孜孜不倦，以專業的態度考慮及綜合了持分者的意見及其他司法管轄區的最佳行事常規。

《模範框架》可說是亞太區首個在 AI 領域針對保障個人資料私隱而制定的指引性框架。人工智能安全乃國家安全的重點領域之一，我相信《模範框架》將孕育 AI 在香港的健康及安全發展，促進香港成為創新科技樞紐，並推動香港以至大灣區的數字經濟發展。

鍾麗玲

個人資料私隱專員

2024 年 6 月

引言

1. 人工智能（AI）沒有通用的定義，泛指一系列模仿人類智能及以電腦程式和機器透過所輸入的數據執行解難、提供建議和預測、作出決策及生成內容等工作（或將其自動化）的科技。

《2021 年 AI 指引》

2. 2021 年 8 月，香港個人資料私隱專員公署（私隱專員公署）出版了*《開發及使用人工智能道德標準指引》*（《2021 年 AI 指引》），主要為開發及使用 AI 系統時涉及使用個人資料的機構提供建議。

3. 《2021 年 AI 指引》建議機構採納三項**數據管理價值**，分別是 (1) 尊重、(2) 互惠，及 (3) 公平。指引亦鼓勵機構採納七項國際認可的**AI 道德原則**，分別是 (1) 問責、(2) 人為監督、(3) 透明度與可解釋性、(4) 數據私隱、(5) 公平、(6) 有益的 AI，及 (7) 可靠、穩健及安全。

圖 1：數據管理價值及 AI 道德原則

	數據管理價值	AI 道德原則
1	尊重	• 問責 • 人為監督 • 透明度與可解釋性 • 數據私隱
2	互惠	• 有益的 AI • 可靠、穩健及安全
3	公平	• 公平

採用 AI 的趨勢

4. 近年來，隨著基礎模型[1]的出現，AI 經歷了翻天覆地的變化。簡而言之，基礎模型是以大量非結構化資料訓練的 AI 模型，能夠廣泛應用於不同的任務、操作和應用程式，且有多種用途。就生成式 AI 而言，基礎模型有多種類型，例如語言模型、音訊模型、視訊模型，甚至多模態模型。例如，大型語言模型是以文字資料訓練的基礎模型，可應用於需要自然語言處理[2]的任務，例如聊天機械人。

5. 儘管小型語言模型日漸增加，但對於許多機構來說，開發大型基礎模型成本高昂亦耗費時間，因此越來越多企業（特別是中小企）在其營運中採用 AI 時，會傾向從銷售商和開發商購買為買家的特定用例而度身訂造的 AI 方案，而不會從零開始開發 AI 系統。如此一來，這些機構能夠使用從 AI 系統開發商及／或銷售商購買的定製 AI 系統或現成方案來加強其決策能力、自動化流程、生成內容，以及從數據中獲取見解。在這種做法下，要達致合乎道德標準地開發及使用 AI，各方需要分別承擔不同責任。

圖 2：AI 模型開發商、銷售商和採購／實施／使用 AI 的機構的概況

1 「基礎模型」一般是指使用廣泛且大規模的數據訓練而成的機器學習模型，其輸出結果具通用性，並且適用於各種下游的特定工作或用途，包括完成簡單的工作、理解自然語言、翻譯及生成內容。

2 根據美國國家標準與技術研究院的說法，自然語言處理是一種強大的計算方法，可讓機器有意義地理解人類的口頭和書面語言。自然語言處理能夠支援演算法搜尋、語音翻譯，甚至對話文字生成等活動，幫助我們與電腦系統溝通以執行各種任務。

本模範框架的重點

6. **本模範框架向採購、實施及使用任何種類的 AI 系統（包括預測式 AI 和生成式 AI）時涉及使用個人資料的機構提供一套建議，旨在為該等機構提供最佳行事常規**。本模範框架建基於一般業務流程，除了支持內地在 2023 年推動的*《全球人工智能治理倡議》*，亦反映國際間認受的規範及最佳行事常規。機構可以採納本模範框架的建議以協助它們遵守一些公認的數據保障原則，當中包括數據安全。由於 AI 使用大量數據，數據安全尤其重要。

7. 在本模範框架中，「機構」是指從第三方採購 AI 方案並收集及使用個人資料以用作 (a) 定製 AI 系統以提升它在特定領域或用例的表現及 / 或 (b) 運作 AI 系統的機構；「AI 供應商」是指向機構提供 AI 方案的 AI 開發商及 / 或 AI 銷售商（視情況而定）。我們建議機構開發自建的 AI 模型時參閱《2021 年 AI 指引》內的建議。

遵從《個人資料（私隱）條例》

8. **機構在採購、實施及使用 AI 方案的過程中處理個人資料時，應確保遵從《個人資料（私隱）條例》（《私隱條例》）的規定，包括附表 1 的六項保障資料原則**。該六項保障資料原則是《私隱條例》的核心規定，涵蓋個人資料由收集到銷毀的整個生命周期。有關保障資料原則的概覽見**附錄 A**。

9. 本模範框架中的建議並非涵蓋所有場景，機構應適當地採取其他措施以遵從《私隱條例》的規定，並且在採購、實施及使用 AI 方案時遵從數據管理價值及 AI 道德原則。

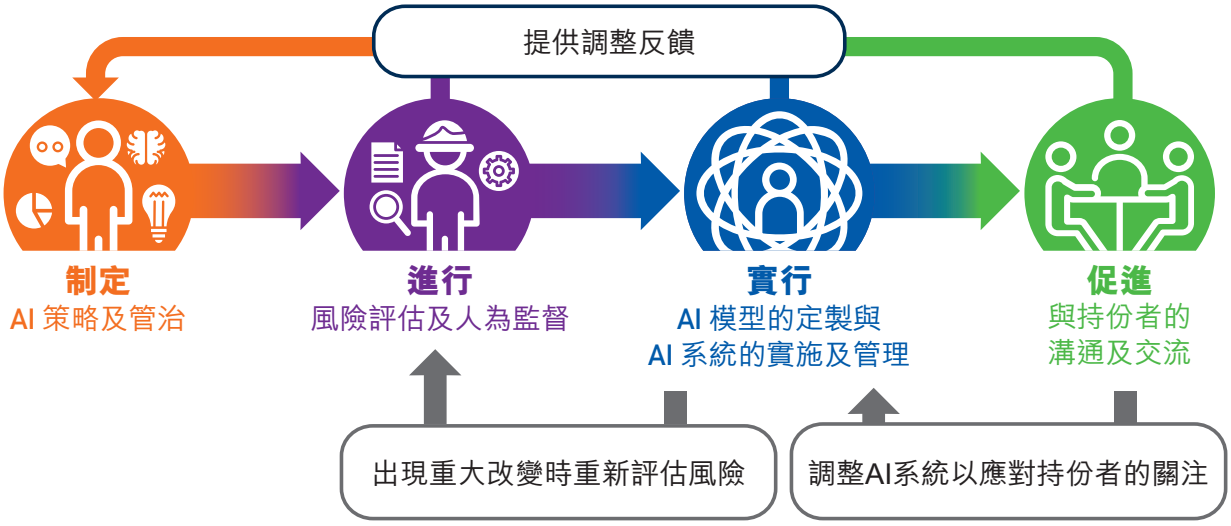
10. 私隱專員公署倡議機構採用個人資料私隱管理系統（Personal Data Privacy Management Programme），以確保機構負責任地收集、持有、處理及使用個人資料，從而加強數據管治。良好的數據管治與管治值得信賴的 AI 息息相關。機構把 AI 管治原則以及「貫徹私隱設計」的精神納入現行的私隱管理系統及 / 或資料管理措施，可鞏固其保障個人資料私隱的決心，並展示奉行問責原則。

個人資料保障模範框架

11. 為確保數據管理價值及AI道德原則（見第3段）得以落實，機構在採購、實施及使用AI方案時，應考慮下述範疇的建議措施，制定適當的政策、措施及程序：

- AI策略及管治（第一部）；
- 風險評估及人為監督（第二部）；
- AI模型的定製與AI系統的實施及管理（第三部）；及
- 與持份者的溝通及交流（第四部）。

圖3：個人資料保障模範框架



12. **一般來說，採購第三方AI方案的機構應以風險為本的方式採購、實施及使用AI系統，作為其更廣泛及全面的AI管治的一部分**。機構在考慮及採用本模範框架的建議措施時，應務求所採取的措施與AI系統在該處境中可能構成的風險相稱。機構可考慮利用及調整現有的數據管治、問責制度及第三方供應商的管理框架，並將此模範框架的元素合併到現有的工作流程中。

個人資料保障模範框架

第一部 AI 策略及管治

13. 高級管理層（如行政層或董事會層）的支持和積極參與是合乎道德標準及負責任地採購、實施及使用 AI 系統的成功要素。**機構應建立內部的 AI 管治策略，一般包含 (i) AI 策略、(ii) 關於採購 AI 方案的管治考慮，以及 (iii) AI 管治委員會（或類似組織），以引領相關過程**。

1.1 AI 策略

主要原則：問責

14. 機構應制定 AI 策略，以展示高級管理層有決心通過合乎道德標準及負責任的方式採購、實施及使用 AI。AI 策略亦應就採購 AI 方案的目的以及如何實施和使用 AI 系統提供相關指引。AI 策略可包含以下要素：

(i) 界定 AI 系統在機構的科技生態系統中提供的功能；

(ii) 參考 AI 道德原則，制定特定適用於機構在採購、實施及使用 AI 方案方面的道德原則；

(iii) 列明 AI 系統在機構中不可接受的用途[3]；

(iv) 建立 AI 清單，以幫助機構實施管治措施；

(v) 就如何合乎道德標準地採購、實施及使用 AI 方案制定具體的內部政策和程序，包括制度化的決策過程和上報準則；

(vi) 確保備有適當的基礎技術設施，以支援合法、負責任和優質的 AI 實施及使用，包括資料儲存、管理和處理工具、運算資源和設施，及用於操作和監察的機器學習營運措施（machine learning operations）等；

(vii) 定期與所有相關人士就 AI 策略、政策和程序溝通，包括各級內部職員和外部持份者（如適當），例如業務夥伴及客戶；

(viii) 考慮可能將會適用於 AI 的採購、實施及使用的法律和法規，包括資料保障及知識產權的法律；及

(ix) 根據有關實施模範框架第二、三和四部的意見，持續檢討和調整 AI 策略。

3 機構應識別潛在風險高至應禁止採用的 AI 用例。這些用例的列表不應固定不變，而應隨著 AI 技術的發展、新風險的出現及 / 或新風險緩減措施的採用而增加、刪除或調整用例。

1.2 關於採購 AI 方案的管治考慮

15. 機構採購 AI 方案時通常涉及**聘請第三方定製 AI 系統或購買 / 訂購現成的 AI 系統 / 服務**，一般涉及以下步驟：

(i) 尋找合適的 AI 方案，並考慮 AI 供應商的專業知識和聲譽；

(ii) 根據機構使用AI的目的，選擇具有合適AI模型的AI方案。考慮的因素包括機器學習演算法的類型（例如迴歸模型（regression models）、決策樹（decision trees）、隨機森林（random forests）、神經網絡（neural networks）等）、學習模型的類型（例如監督式學習模型、非監督式學習模型和強化學習模型）及模型的規模和複雜性；

(iii) 收集和準備機構的數據，以定製 AI 模型（如有需要）；

(iv) 為特定目的定製 AI 模型（如有需要）；

(v) 測試、評估和驗證 AI 模型；

(vi) 測試和審核系統和組件的安全性和私隱風險；及

(vii) 將 AI 方案納入機構的系統中。

圖 4：採購及實施 AI 模型的過程

16. 我們建議打算投資購入 AI 方案的機構考慮以下管治因素：

(i) 使用 AI 的目的，及預計使用 AI 的用例；

(ii) 向準 AI 供應商提出有關私隱和保安的主要責任及道德規定[4]；

(iii) 準 AI 供應商應遵循的技術性和管治方面的國際標準[5]；

(iv) AI 方案在甚麼準則下（例如通過評分的方式）須交由 AI 管治委員會（或類似組織）審查（例如 AI 的用例可能導致高風險的情況，見第 2.1 節），以及相關程序；

(v) 如機構採購 AI 方案涉及聘用資料處理者（例如直接在第三方平台上開發或定製 AI 模型，及／或 AI 方案在「AI 即服務」（AI-as-a-service）的雲端平台上運作[6]），所需要簽署的資料處理者協議；

(vi) 處理 AI 系統生成結果的政策（例如，在可行的情況下，採用技術將 AI 生成內容中的個人資料匿名化，將 AI 生成內容添加標記或水印，並過濾可能引起道德問題的 AI 生成內容）；

(vii) 持續分析業務和技術環境的計劃，以找出潛在研究或策略，從而幫助機構將「貫徹私隱設計」及「貫徹道德設計」的原則納入 AI 管治中；

(viii) 持續監察、管理和維持 AI 方案的計劃（見第 3.3 節），並尋求 AI 供應商的協助（如適當）；及

(ix) 在盡職調查過程中，就 AI 供應商的能力進行評估。

4 連同其他事宜，該等責任及規定應與機構的私隱政策（須遵從《私隱條例》）和 AI 道德原則保持一致，例如，根據用例和情況，該等責任及規定可應對數據集的公平性、適合機構目的的機器學習演算法種類和學習類型，以及如何符合道德期望（例如，不同 AI 模型有不同透明度和可解釋性的考慮，見第 2.3 節）。

5 機構可參考國際標準化組織（ISO）和電機電子工程師學會（IEEE）等專業協會制定和發布的標準。例如，ISO/IEC 27001:2022 和 ISO/IEC 27002:2022 涵蓋資訊安全，ISO/IEC 27701:2019 涵蓋個人資料保護，ISO/IEC 23894:2023 涵蓋 AI 的風險管理，和 ISO/IEC 42001:2023 涵蓋機構內 AI 管理系統的建立、實施、維護和持續改進。

6 私隱專員公署鼓勵機構參閱其有關外判個人資料的處理的資料單張以獲取更多資訊：https://www.pcpd.org.hk/tc_chi/publications/files/dataprocessors_c.pdf。

圖 5：採購 AI 方案的管治考慮

	使用 AI 的目的
	私隱和保安的責任及道德規定
	技術性和管治方面的國際標準
	審查 AI 方案的準則和程序
	資料處理者協議
	處理 AI 系統生成結果的政策
	持續檢視環境變化的計劃
	持續監察、管理和維持 AI 方案的計劃
	評估 AI 供應商

17. 在採購及實施 AI 模型的每一個階段中（見圖 4），機構的參與程度或會有所不同，這取決於機構為開發及／或定製 AI 模型所提供的數據及指示等。以下列情況為例，機構的參與程度在各情況中皆有不同：

- 由第三方開發商開發的全定製 AI 模型；
- 根據機構的需求對預先訓練的 AI 模型進行輕微定製的 AI 模型；
- 現成的 AI 方案，包括「AI 即服務」（AI-as-a-service）及在雲端上運作的服務（如透過應用程式介面（API）使用的服務）；或
- 使用第三方的自動機器學習服務平台，由機構提供數據或定製指示以建立的 AI 模型。

18. 在每種情況下，機構聘請第三方均可能引起與資料（包括個人資料）保障相關的合規事宜。機構與第三方都應在雙方簽署的服務協議中，明確處理有關事宜。

圖 6： 主要的資料（包括個人資料）保障合規事宜

誰是資料使用者？

- 控制個人資料的收集、持有、處理或使用的一方屬資料使用者（《私隱條例》第 2 條）。
- 例如，決定使用哪些類型的個人資料來定製、測試、驗證以及／或操作 AI 系統的機構很可能被視為資料使用者。

誰是資料處理者？

- 代他人處理個人資料且不為自身目的處理該資料的一方屬資料處理者（《私隱條例》第 2 條）。
- 例如，如 AI 供應商在為定製 AI 處理個人資料時不就 AI 模型的輸入數據和輸出的結果作出決定，且只為 AI 的訓練／定製過程提供平台，便可能是資料處理者。

跨境轉移資料的合法性

- 如在雲端平台上定製及使用 AI 涉及處理個人資料，而該平台的數據中心分佈於多個司法管轄區，當機構（作為資料使用者）將個人資料轉移至香港以外地方時：
  - 必須遵守《私隱條例》的相關規定，包括六項保障資料原則；及
  - 應查明是否有跨境資料轉移的法例限制或規管個人資料由處理資料的司法管轄區轉回至資料使用者。

資料保安的考量

- 如機構作為資料使用者為定製及 / 或使用 AI 將個人資料轉移予資料處理者，則必須採用合約規範或其他方法，按照保障資料第 4(2) 原則，防止個人資料未獲准許或意外地被查閱、處理、刪除、喪失或使用。

19. 採購團隊應與項目團隊共同選擇合適的 AI 方案、決定機構的參與程度，以達到機構的目的[7]，**並與法律和合規部門合作，以應對資料保障方面的合規問題**。

1.3 管治架構

主要原則：問責 / 人為監督

20. 我們建議採購、實施及使用 AI 系統的機構具備不同領域的專業知識，例如電腦工程、數據科學、網絡安全、用戶體驗設計、法律與合規、公共關係等。**機構應建立具足夠資源、專業知識和決策權的內部管治架構，以引領 AI 策略的實施，並監督 AI 系統的採購、實施及使用**。AI 的管治架構可包含以下要素：

(i) AI 管治委員會（或類似組織）監督所有 AI 方案的整個生命周期（由採購、實施、使用以至終止），並向董事會匯報；AI 管治委員會的監督應橫跨整個業務，而不受部門劃分（即風險和合規、財務或銷售等）的限制。

7 例如，採購團隊應考慮 AI 系統的輸出結果所需的準確性和可解釋性程度，以及將系統實施至機構的 IT 基礎設施中的障礙。

AI 管治委員會

高級管理層參與及跨專業領域合作應是 AI 管治委員會最重要的特質。機構應成立包含不同技能和觀點的跨部門團隊，包括業務運作人員、採購團隊、系統分析師、系統架構師、數據科學家、網絡安全專家、法律及合規專業人員（包括保障資料主任）、內部審計人員、人力資源人員、客戶服務人員等。

機構應指派高級管理人員（例如行政總裁、資訊總監 / 技術總監、私隱總監或來自高層的類似職位）領導該跨部門團隊。

（選擇性措施）AI 管治委員會可向外部專家尋求 AI 及道德標準方面的獨立意見。如某項目規模龐大、影響廣泛及 / 或備受注目，其道德價值有可能受到挑戰，機構可另外成立 AI 道德委員會以進行獨立檢視。

(ii) 為不同部門或人員訂明清晰的角色及責任；

角色及責任的例子：

- 採購團隊應根據機構 AI 策略中規定的內部政策和程序購買 AI 方案；
- 系統分析師、系統架構師及數據科學家應專注於 AI 方案的定製、實施、監察及維護，以及機構的內部數據管治流程；
- 法律及合規專業人員應專注於確保機構遵從與 AI 系統的採購、實施及使用相關的法律及規例（包括資料保障法律）以及內部政策；
- 審查員應專注於審查 AI 系統的決策和輸出的結果；
- 業務及運作人員應按機構的政策和程序使用 AI；及
- 客戶服務及公關人員應與持份者（包括顧客、監管機構和公眾）溝通並回應其關注。

(iii) 在財政和人力上有足夠的資源；及

需要足夠資源（如具備相關技能、經驗及專門知識的專家）的例子：

- 在有需要時進行風險評估，以識別及減低使用 AI 所帶來的風險，包括私隱、保安及道德風險，並採取相應的措施減低風險；
- 建立能夠協助機構監察、記錄及檢視已實施的 AI 方案的內部數據管治流程及資訊系統；及
- 為相關人員提供足夠的培訓（見第 1.4 節）。

(iv) 建立有效的內部匯報機制，用於匯報任何的系統故障或提出有關資料保障或道德問題，以便 AI 管治委員會作出恰當的監察。

機構應制定 AI 策略及成立 AI 管治委員會（或類似組織），以引領 AI 系統的採購、實施及使用。

1.4 培訓及加強認識

主要原則：問責

21. 為確保有關 AI 的政策得以施行，機構應為所有相關人員提供足夠的培訓，以確保他們具有適當的知識、技能和認識，以便在使用 AI 系統的環境中工作。

圖 7： ***培訓的例子***

建議人員	培訓主題
系統分析師 / 系統架構師 / 數據科學家	• 遵從資料保障法律、規例，和內部政策；網絡保安風險
AI 系統使用者（包括業務運作人員）	• 遵從資料保障法律、規例，和內部政策；網絡保安風險；一般 AI 科技
法律及合規專業人員	• 一般 AI 科技和管治
採購人員	• 一般 AI 科技和管治
審查員	• 查找並糾正 AI 系統所作的決定或所產生的內容中任何不義的偏見、非法的歧視和錯誤 / 不準確之處
所有工作上與 AI 系統有關的人員	• 機構所使用的 AI 系統的好處、風險、功能和限制

審查員的角色

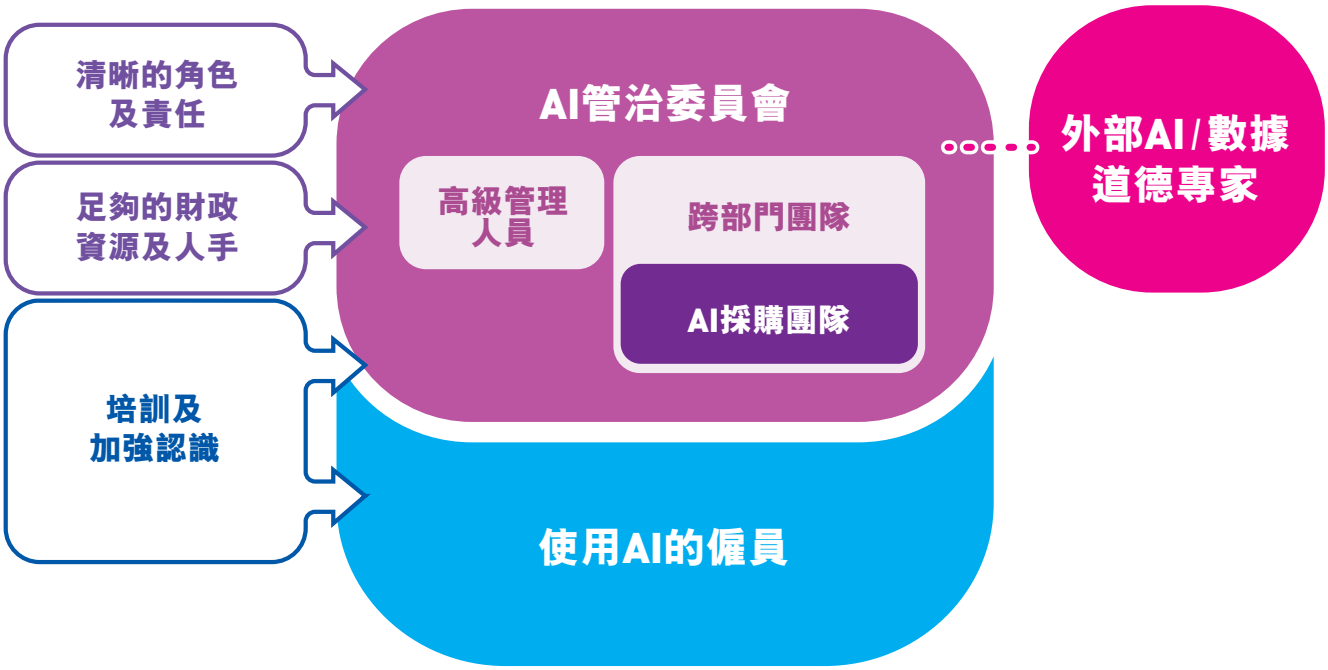
為確保審查員認真履行其職責，以及人為監督並非只屬象徵性質，相關人員應有能力評估及解釋 AI 所作的建議及 / 或審查 AI 生成內容。審查員應能恰當地行使酌情權和權力，在有需要時否決 AI 所作的建議或標記有問題的建議，並提醒 AI 供應商。

在適當的情況下，機構可考慮要求 AI 供應商提供有關 AI 輸出的結果的資訊和解釋，以便有效行使人為監督。

22. 作為私隱管理系統的一部分，任何有關個人資料私隱保障的培訓（涵蓋《私隱條例》的規定及機構的私隱政策）亦應涵蓋採購、實施及使用 AI 系統時收集及使用個人資料的情況。

23. 此外，機構亦應通過員工會議或其他內部溝通渠道向所有相關人員傳達合乎道德的 AI 的重要性和所適用的原則，以培養和促進合乎道德和保障私隱的文化。

圖 8： ***管治架構***



第二部 風險評估及人為監督

24. 機構採購並應用在機構營運的特定用途上的 AI 方案，可能包含原為一般或多項用途而開發的 AI 模型。因此，AI 系統的風險程度取決於機構如何使用 AI 系統及使用該系統的特定目的。例如：

- 用作評估個人信貸的 AI 系統的風險一般比用於推送個人化廣告的 AI 系統的風險較高，因為前者或會令個人無法獲得信貸安排，而後者不太可能對個人造成重大影響。
- 用於內部翻譯的生成式 AI 工具與直接回覆客戶的生成式 AI 聊天機械人相比，對個人產生重大影響的可能性較小。
- 在決策上有完全自主能力的 AI 系統亦可能比涉及一定程度人為操作的 AI 系統（例如只向人類決策者提供建議的 AI 系統）有較高風險，特別是可能對個人造成重大影響的決策。

25. **在採購、使用及管理 AI 系統時，應採取風險為本的方式。機構需要進行全面的風險評估，有系統地識別、分析及評估過程中涉及的風險，包括私隱風險。機構應建立一套風險管理機制，在 AI 系統的整個生命周期內持續實施及維持該機制，並將相關資料記錄存檔**[8]。機構應禁止在機構的 AI 策略（見第 1.1 節）中被評定為帶有不可接受的風險的 AI 用例。

機構需要進行全面的風險評估，有系統地識別、分析及評估採購、使用及管理 AI 過程中涉及的風險，包括私隱風險。

26. **在採購過程時或對現有的 AI 系統進行重大更新時，應由跨部門團隊進行風險評估。為了識別與私隱相關的風險（例如通過私隱影響評估），跨部門團隊應包含私隱合規人員**。視乎有關情況，機構可能需要諮詢來自不同社會、文化和宗教背景，以及不同性別及種族的人士（或具有相關知識的專家），以便識別使用 AI 時潛在不義的偏見和非法的歧視、對個人權利、自由和利益的不利影響，以及更廣泛的社會影響。**所有風險評估應妥善記錄存檔，而結果應按照機構的 AI 管治委員會認可的 AI 政策進行檢視**。

8 AI 管治委員會可考慮參考 ISO/IEC 23894:2023（資訊科技 - 人工智能 - 風險管理指南）及美國國家標準暨技術研究院的 AI 風險管理框架，將風險管理整合到 AI 系統的生命周期中。

圖 9： ***風險評估的程序***

2.1 須考慮的風險因素

主要原則：有益的 AI ／數據私隱／公平

27. **由於使用 AI 往往涉及使用個人資料，因此必須應對資料私隱風險**。為保障個人資料私隱，機構在進行風險評估時應考慮以下因素：

(i) 用來定製所採購 AI 方案的資料及／或輸入 AI 系統用作決策的資料的准許用途，當中須顧及《私隱條例》下的保障資料第 3 原則[9]；

(ii) 個人資料的數量（當中須顧及《私隱條例》下的保障資料第 1 原則）[10]：

- 定製 AI 模型所須的個人資料的數量；
- AI 系統在運作中收集的個人資料的數量（例如監控、系統性的評估和監察可能涉及大規模收集個人資料）；及
- AI 供應商在開發和訓練 AI 方案時需要的個人資料的數量，以及 AI 供應商有否盡可能採用匿名化技術，以遵守資料最少化的原則；

9 保障資料第 3 原則訂明，未得資料當事人的訂明同意，個人資料不得用於新目的。
10 保障資料第 1 原則訂明，所收集的個人資料就收集目的而言，屬足夠但不超乎適度。

(iii) 所涉及資料的敏感程度[11]，當中須顧及《私隱條例》下的保障資料第 4 原則[12]；

(iv) 所涉及資料的質素，當中須考慮其來源、可靠性、真實性、準確性（當中須顧及《私隱條例》保障資料第 2 原則）、一致性、完整性、相關性及可用性[13]；

(v) 在使用 AI 系統時的個人資料保安[14]，當中應考慮如何在機構的科技生態系統以及 AI 系統之間轉移個人資料[15]，以及有否對 AI 生成的結果採取保護措施，以減輕個人資料外洩的風險，當中須顧及《私隱條例》保障資料第 4 原則[16]；及

(vi) 私隱風險（例如個人資料的過度收集、濫用或外洩）出現的可能性及其潛在損害的嚴重程度。

28. 從更廣泛的道德標準角度來看，如使用 AI 系統可能會對持份者（尤其是個人）的權利、自由或利益造成影響，風險評估亦須考慮：

(i) AI 系統對受影響個人、機構及社會大眾的潛在影響（包括益處和損害）；

(ii) AI 系統對個人的影響出現的可能性，以及其嚴重程度和持續時間[17]；及

(iii) 降低風險的緩減措施（包括技術性與非技術性措施）是否足夠（見第 2.2 節及第三部）。

29. 使用 AI 系統對個人的潛在影響或會牽涉其法律權利、人權（包括私隱權）、就業或教育前途，以及獲得服務的機會和資格等。**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。**

11 一般被視為較敏感的個人資料包括生物辨識資料、健康資料、財務資料、位置資料、有關受保護特徵的個人資料（例如性別、種族、性取向、宗教信仰、政治立場等），以及弱勢群體（例如兒童）的個人資料。

12 保障資料第 4(1)(a) 原則訂明資料使用者須採取所有切實可行的步驟，以確保由資料使用者持有的個人資料（包括採用不能切實可行地予以查閱或處理的形式的資料）受保障而不受未獲准許的或意外的查閱、處理、刪除、喪失或使用所影響，尤其須考慮該資料的種類及如該等事情發生便能造成的損害。

13 保障資料第 2 原則規定，資料使用者須採取所有切實可行的步驟，以確保在顧及有關的個人資料被使用於的目的下，該個人資料是準確的。

14 機構在使用第三方建立或維護的 AI 方案時需要謹慎評估安全風險，因為 AI 方案可能同時依賴內部開發及 / 或基於公開原始碼和框架的各種形式的軟件和硬件（見第 3.2 節）。

15 保障資料第 4 原則規定，資料使用者須採取所有切實可行的步驟，以確保由其持有的個人資料受保障。

16 保障資料第 4(1)(e) 原則訂明資料使用者須採取所有切實可行的步驟，以確保由資料使用者持有的個人資料（包括採用不能切實可行地予以查閱或處理的形式的資料）受保障而不受未獲准許的或意外的查閱、處理、刪除、喪失或使用所影響，尤其須考慮為確保在保安良好的情況下傳送該資料而採取的措施。

17 例如，機構應考慮 AI 的自主程度、與環境直接互動的能力、該環境的複雜性，以及 AI 所作的決策的複雜性。

當評估 AI 系統對個人造成的潛在影響是否重大時，機構可先考慮其使用可能對個人造成甚麼類型的損害[18]。

若使用 AI 系統產生的輸出結果（例如對求職者的評估）在某些情況下很可能對個人造成嚴重及長期的損害，且該風險無法充分減低，有關 AI 系統應被視為高風險。

圖 10：AI 系統風險評估須考慮的因素（參考例子）

《私隱條例》的規定	資料的數量、敏感程度及質素
資料保安	對個人、機構及社會的潛在影響
影響出現的可能性、嚴重程度和持續時間	緩減措施

2.2 決定人為監督的程度

主要原則：人為監督

30. 風險評估的主要目的是讓機構識別潛在的風險，並採取適當的風險緩減及管理措施。機構應採取風險為本的方式，所採取的緩減風險措施的類別和程度，應與已識別的風險程度相符和相稱。機構應把無法消除的剩餘風險告知 AI 系統的終端使用者及／或受 AI 系統影響的個人。無論如何，剩餘風險應盡量降至可接受的水平。如剩餘風險已在合理地切實可行的情況下減至最低，且 AI 系統可為持份者帶來的益處遠超過其所帶來的風險，則被視為可接受。

在採取風險為本的方式時，所採取的緩減風險措施的類別和程度應與已識別的風險程度相符和相稱。

18 例如，機構應考慮財務損失、身體傷害、歧視、個人資料控制權的喪失、自主權的缺乏、心理傷害，以及其他對權利和自由的不利影響。

31. **人為監督是減低使用 AI 的風險的主要措施**。使用 AI 系統時所需的適當人為監督的程度應基於風險評估的結果。無論如何，人類決策者應就 AI 所作的決策及輸出的結果負上最終的責任。

32. 一般來說，風險較高（即較有可能對個人造成重大影響）的 AI 系統，須有較高程度的人為監督。因此：

(i) 高風險的 AI 系統應採取「人在環中」（human-in-the-loop）方式進行人為監督。在這模式中，人類決策者在決策過程中保留著控制權，以防止及／或減低 AI 出錯或輸出不當的結果及／或作出不當的決定。

(ii) 風險最小或較低的 AI 系統可採取「人在環外」（human-out-of-the-loop）方式進行人為監督。在這模式中，AI 系統可在沒有人為介入下採納輸出的結果及／或作出決定，以達致完全自動化／完全自動化的決策過程。

(iii) 如「人在環中」及「人在環外」兩種方式均不適合，例如當風險程度不可忽視，或「人在環中」的方式不能符合成本效益或不可行，機構可考慮「人為管控」（human-in-command）方式。在這模式中，人類決策者會利用 AI 系統輸出的結果，監督 AI 系統的運作，並在有需要時才介入。

圖 11： *風險為本的人為監督*

圖 12： ***可能帶來較高風險的 AI 應用例子***

- 使用生物識辨資料實時識別個人
- 評估個人享用社會福利或公共服務的資格
- 求職者評估、工作表現評核或終止僱傭合約
- 評估個人的信用可靠程度，以作出自動化決策
- AI 輔助醫學影像分析或治療

33. **機構使用 AI 系統時，應視乎其風險程度，進行適當的人為監督**。我們建議機構從供應商處了解 AI 模型的開發及訓練過程中有否及如何由審查員進行監督，以降低使用 AI 模型時對個人造成重大不利影響的風險。機構也可能需要要求 AI 供應商就 AI 輸出的結果提供資訊和解釋，以在使用 AI 系統時有效地執行人為監督。

2.3 減低風險的權衡

34. 機構試圖減低 AI 的風險以符合 AI 道德原則時，可能會遇到一些互相矛盾的情況，需要作出平衡，並作出取捨（見圖 13）。

35. **機構可能要考慮會使用 AI 作決策或生成內容的情況，以決定如何合理地作出權衡**。例如，若 AI 系統的決策會影響客戶能否享用服務，而提供人為監督的審查員亦需要向客戶解釋 AI 的決策，AI 模型的可解釋性便可能比輸出結果的準確性更為重要。**我們建議機構記錄權衡時作出的評估，包括作出最終決定的理由**。

36. 無論如何，機構都必須遵守任何適用的法律要求，包括《私隱條例》的要求。

圖 13： 減低風險時需要作出權衡的例子

預測準確性 / 表現

一些AI模型（例如決策樹）較容易解釋，但其預測準確性較低

1

輸出結果的可解釋性

深度學習神經網絡模型的預測結果一般較準確，但被認為是較難解釋的「黑盒」

資料統計準確度

2

機構可能需要使用更多資料（包括個人資料）訓練、定製及/ 或測試AI模型以提高其準確性和公平性

資料最少化

機構應注意只使用就其目的而言足夠、必要但不超乎適度的個人資料

可解釋性

3

為AI系統的輸出結果提供解釋會提高可解釋性

數據安全 / 私隱

如提供詳細資訊，可能會揭露有關AI模型的內部運作原理，增加其受到攻擊和發生資料保安事故的機率

私隱增強技術

4

使用私隱增強技術（例如合成數據[19] 和差分私隱技術[20]）可以減少所使用的個人資料

輸出結果的準確性

機構應留意私隱增強技術可能會影響AI輸出結果的準確性

19 合成數據（synthetic data）是指人工生成的數據集，與真實人士無關。

20 差分私隱（differential privacy）是在發放數據集時確保私隱受到保障的方式，做法通常是在發放數據集前在當中加上雜訊（即作出輕微的改動）。與去識別化不同，差分私隱不是一個特定的程序，而是通過某些程序後可令數據集達致的質素或狀態。如果不能確定個別人士的資料是否包含在一個已發放的數據集中，該數據集便達致差分私隱的狀態。差分私隱對私隱提供的保障一般被視為較去識別化強。

第三部　AI 模型的定製與 AI 系統的實施及管理

37. 除了採購合適的 AI 方案以達到機構的目的外，用於定製 AI 模型及使用 AI 系統的數據數量和質素亦會對 AI 系統的可用性、準確性和可靠性有重大影響。定製 AI 方案的主要目的是利用數據提供更多有關特定領域 / 情況的資訊以提升 AI 方案的表現。在本模範框架中，「定製」（customisation）是指為符合機構使用 AI 的目的，而調整或改造預先訓練的 AI 模型的過程，包括 AI 模型的微調 [21] 及接地 [22]。

38. AI 模型或會持續學習和演變，而 AI 系統的操作環境亦可能會轉變。因此，我們建議機構在採用 AI 模型後持續監察和檢視情況，並向使用者提供支援，以確保 AI 系統保持有效、相關和可靠。

圖 14：定製與管理的主要流程

21 微調（fine-tuning）是採用以大型和通用數據集訓練的 AI 模型，並使用其他特定數據對其進行更新 / 調整以實現特定目的或需求的過程。

22 接地（grounding）是將 AI 模型與可驗證的現實世界的知識和外部來源的範例聯繫起來的過程。就生成式 AI 模型而言，最流行的接地方法之一是檢索增強生成（retrieval-augmented generation）。它透過加入一個提供接地數據的資訊檢索系統來增強大型語言模型能力，以提高該大型語言模型在特定用例或領域的表現。

3.1 為定製及使用 AI 準備數據

主要原則：數據私隱 / 公平

39. 機構或會於 AI 定製和決策或輸出結果的階段使用內部專有數據，該些數據往往涉及個人資料。**定製及操作 AI 時，良好的數據管治不單保障個人資料私隱，亦可確保數據質量，這對 AI 系統的穩健性和公平性至為重要**。管理不善的數據或會引致「廢料進，廢品出」的情況，並可能對 AI 系統產生的結果有不利影響（例如預測式 AI 輸出不公平的結果、生成式 AI 出現「幻覺」等情況[23]）。

良好的數據管治對 AI 系統的穩健性和公平性至為重要。

40. 機構應適當考慮其上游 AI 供應商的數據管治措施以及訓練數據的來源。訓練數據的合法性和質素或會影響 AI 方案的質素、穩健性和公平性，以及 AI 方案是否符合相關法律要求。

41. 機構在準備用於 AI 定製和使用的數據集時應採取以下步驟：

(i) **必須採取措施，確保遵從《私隱條例》的規定**，包括：

- 以合法及公平的方法收集足夠但不超乎適度的個人資料 - 見保障資料第 1 原則；
- 除非已取得資料當事人的明確及自願的同意，或有關個人資料已被匿名化，否則避免將個人資料用於與原本的收集目的不相符的目的 - 見保障資料第 3 原則；
- 在使用個人資料前，採取所有切實可行的步驟，確保有關資料準確 - 見保障資料第 2(1) 原則；
- 採取所有切實可行的步驟，確保個人資料安全 - 見保障資料第 4 原則；
- 在達致原本的收集目的後，刪除有關個人資料或將資料匿名化 - 見《私隱條例》第 26 條及保障資料第 2(2) 原則；

23 不善的數據管治未必是生成式 AI 出現「幻覺」的唯一原因。使用轉換器架構（transformer architecture）模型的生成式 AI 往往會帶有「幻覺」（hallucination）的情況，但通過接地及提示工程（prompt engineering）等方式可有效減低「幻覺」的出現。

- 在收集個人資料之時或之前，採取所有切實可行的步驟，確保資料當事人獲告知所需資訊，例如資料可能移轉予甚麼類別的人，尤其是當涉及 AI 供應商時 - 見保障資料第 1 原則；
- 採取所有切實可行的步驟，確保有關機構的個人資料政策和措施的必要資訊（例如透過私隱政策提供）可供查閱 - 見保障資料第 5 原則；及
- 實施可協助機構依從資料當事人的要求的系統 - 見保障資料第 6 原則。

(ii) **在定製及使用 AI 模型時盡量減少所使用的個人資料數量，從而減低私隱風險（保障資料第 1 原則）。同時，機構亦必須收集足夠的數據以確保結果是準確和不存在偏見**。機構應採取以下措施和技術（如適用）：

- 只收集及使用為達致特定目的而定製及／或操作 AI 相關的個人資料，刪除含有個人特徵但與該目的無關的資料，即使獲取更廣泛的數據集可以令訓練模型輸出更準確和更公平的結果。機構可預先與 AI 供應商諮詢數據科學家及主題／領域專家，以識別定製 AI 模型所需和足夠的數據量；

例子 1：一家時裝零售平台正計劃採購第三方開發的 AI 聊天機械人，並將其進行定製，以為客戶推薦時裝建議。該公司或認為需使用不同客戶群過去的購買記錄和瀏覽紀錄來微調聊天機械人。然而，客戶的姓名、聯絡資料、某些人口特徵等個人資料的使用並非是需要的。

- 考慮 AI 模型的規模和複雜性。如不需要採用複雜或定製的模型已能達致預期目的，則選擇需要較少數據來定製的較簡單和較小模型，或選擇不需要數據來定製的現成模型；
- 在適當情況下使用匿名化[24]、假名化[25]或合成數據定製及輸入 AI 模型[26]；

24 匿名化數據（anonymised data）是指經過處理從而不能從中識別個人身分的數據集。由於匿名化數據不能用來識別個人，因此不是個人資料。

25 在假名化數據（pseudonymised data）內，可識別個人身分的資料已被移除，並由其他數值取代，以防止在沒有額外資料下直接從有關數據集識別出個人的身分。假名化數據仍屬個人資料，因為在額外資料的輔助下，個人仍然可被間接地識別出來。

26 這三種資料最少化技術可能不適用於某些類型的非文字數據，例如圖像。

- 在發放數據集供定製 AI 模型使用之前，對數據集應用「差分私隱」等私隱增強技術（PETs）；
- 將不再需要用於 AI 定製及使用的個人資料[27]從 AI 系統刪除；及
- 如採購專家系統的 AI 模型[28]足以達致相同的目的而無需要使用大量數據定製 AI，則重新考慮使用個人資料的必要性。

(iii) **管理用以定製及使用 AI 模型的數據質素（保障資料第 2 原則）**，尤其是用於高風險 AI 系統的數據。有關數據應該是可靠、準確、完整、相關、以合法的方式獲取[29]和能夠代表目標群體，並且就定製 AI 模型的目的而言，沒有不義的偏見和非法的歧視。因此，機構應考慮以下事宜：

- 了解用於定製模型的數據的來源、準確性、可靠性、真實性、一致性、完整性、相關性及可用性；
- 進行相關的數據準備程序，例如注解、標籤、清理、補充及聚合；
- 識別數據集內的離群值和異常情況，如有需要時，移除或取代有關數值，並保留相關變更的紀錄；
- 在使用數據定製 AI 模型前，測試有關數據的公平性；

例子 2：如某些組別人士的代表人數不足或超出比例，便可能令數據集本身存在不義的偏見。要處理此問題，可採用不同的抽樣方法來重新平衡各類別的樣本的分佈。例子包括隨機增加少數法（random over-sampling，即複製少數類別的樣本）及隨機減少多數法（random under-sampling，即刪除多數類別的樣本）[30]。

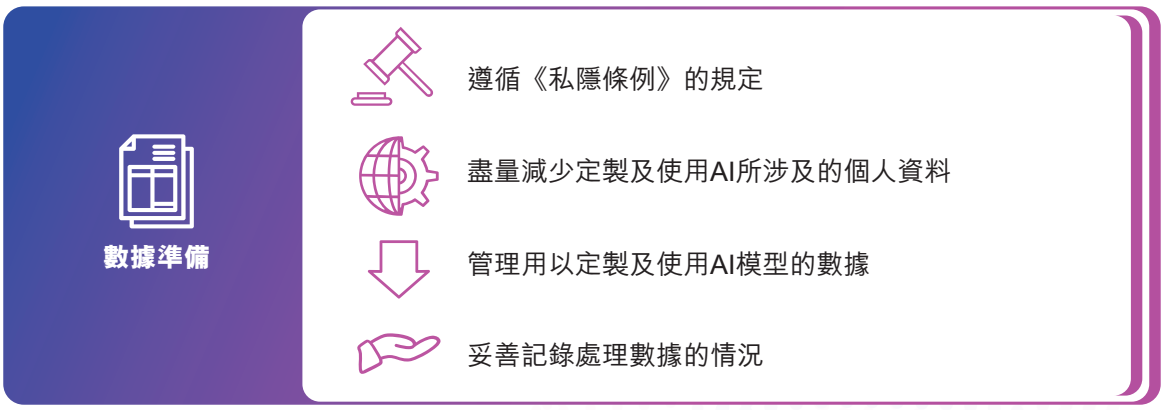
27 例如，如個人資料在接地過程中因應個別要求被輸入生成式 AI 系統，該個人資料應在相關要求完成後刪除。
28 專家系統（expert system）是「從知識庫作出推斷，以複製特定領域內人類專家的決策能力的一種 AI」（來源：IAPP AI 術語表，原文為英文）。專家系統可透過根據該領域的專門知識創建的規則來構建，而無需依賴數據和機器學習。
29 例如，根據《私隱條例》，個人資料應以合法的方法收集。機構還應考慮其他適用的法律，包括知識產權的法律。
30 其他減輕潛在偏見的方法包括重新權衡（re-weighing）輸入神經網絡（neural network）的數據 / 特徵，及除去可能產生偏見的特徵（例如種族）及其代理物的影響。

- 在數據集預留部分數據作「保留」數據（hold-out data）／測試數據（test data）[31]，用於定製 AI 模型之後作驗證和測試之用；及
- 指派人員定期檢視是否需要用更多數據進一步定製 AI 模型，以確保模型的有效性。

(iv) **應妥善記錄為定製及使用 AI 而處理數據的情況**，以確保長期維持數據的質素和安全，以及符合《私隱條例》的規定。應記錄的資料包括：

- 數據的來源；
- 數據的准許用途；
- 所用的數據是如何從可供使用的數據中揀選出來；
- 數據是如何收集、篩選、在機構內轉移及從機構轉移至 AI 供應商（如適用）；
- 數據儲存的位置；及
- 如何長期維持數據質素。

圖 15：數據準備的四個範疇



31 在監督學習中，保留數據集（hold-out dataset）可靠地代表訓練／微調數據集，但由於 AI 模型之前未見過該數據集，因此可用於測試模型能否能有效應用於原來的訓練數據集以外的地方。

3.2 AI 方案的定製及實施

主要原則：透明度與可解釋性 / 可靠、穩健及安全

定製、測試和驗證

42. 如需定製已採購的 AI 模型，機構便需要使用已準備的數據來進行定製，以符合機構的特定需求和使用 AI 的目的 [32]。

43. **機構應按照所涉的風險程度，特別在 AI 模型經定製的情況下，對 AI 模型進行嚴格測試及驗證，以確保有關模型按預期運作，並在使用前評估其可靠性、穩健性和公平性**。建議採取的措施包括：

(i) 驗證 AI 系統是否符合向 AI 供應商說明的私隱責任和道德要求（包括公平性、透明度和可解釋性）；

(ii) 測試 AI 模型是否有錯誤 [33]，以確保其可靠性、穩健性及公平性。有需要時，可諮詢系統分析師、系統架構師和數據科學家以作測試，例如：

- 將 AI 的決策與由人類或傳統非 AI 模型所作的決策互相比較，並將 AI 生成內容與現實世界的數據進行比較；
- 以適當的公平性指標 [34] 及準確性指標 [35] 測試有關 AI 模型的公平性及準確性；
- 使用「保留」數據 / 測試數據測試已定製的 AI 模型，以確保 AI 模型不會過度擬合其訓練數據集 / 定製數據集，並有效運作 [36]；
- 使用邊緣案例及可能出現的惡意輸入，測試 AI 模型；及
- 對 AI 系統進行可重複性及可再現性 [37] 的測試；

32 例如，處理內部文件和數據，協助草擬特定專業領域的文件，或以特定的企業風格生成內容。

33 例如，迴歸測試（regression testing）（即進行測試以確認最近的代碼 / 程式的更改不會對現有 AI 應用程式的表現產生負面影響）。

34 在分類模型（classification model）中，公平性可通過不同的指標（群體均等（demographic parity）、機會平等（equality of opportunity）等）在數學上定義。某些公平性指標不能相互兼容，無法同時滿足。機構應根據具體情況選擇合適的指標使用。

35 準確性可通過不同的指標（例如準確度（accuracy）、精確度（precision）、召回率（recall）、F1 分數（F1 score）、特異度（specificity））在數學上定義，這些指標可測試分類模型中不同類型的錯誤。某些準確性指標不能相互兼容，無法同時滿足。機構應根據具體情況選擇合適的準確性指標使用，以了解需要優化的地方。

36 過度擬合（overfitting）是指「AI 模型變得過於針對訓練數據，無法泛化到未見過的數據，這意味著它可能無法對新的數據集做出準確的預測」（來源：IAPP AI 術語表，原文為英文）。過度擬合亦一般會使 AI 系統更容易受到攻擊，可能會危及訓練 / 定製數據集內個人資料。

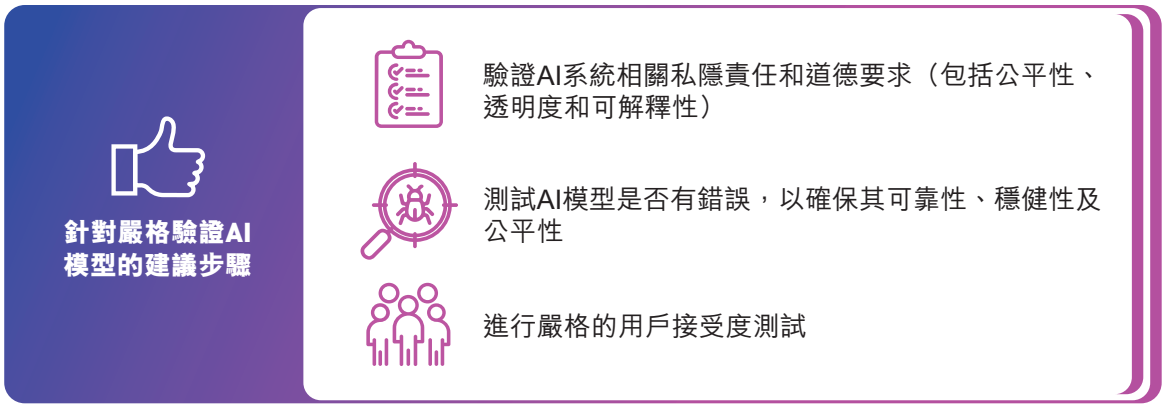
37 可再現性指當使用相同的數據集或預測方法時，AI 系統是否產生相同的結果。可再現性對評估 AI 系統的可靠性十分重要。

(iii) 對 AI 生成內容實施機制以：

- 確保任何個人資料的披露均符合《私隱條例》的規定（如適用）；
- 在可行和適當的情況下，識別內容的生成性質（例如通過標籤和水印）；及
- 在可行和適當的情況下，過濾可能引起道德問題的內容（例如有偏見的輸出、有害內容）；及

(iv) 應在將 AI 方案與機構系統整合之前，進行嚴格的用戶接受度測試。

圖 16：定製、測試及驗證 AI 方案



整合及託管

44. 視乎機構如何將 AI 方案整合（即在機構的本地內部伺服器或由第三方提供的雲端伺服器運行），機構或需考慮其他因素，以符合《私隱條例》的規定。將 AI 系統託管在機構的內部伺服器，與託管在第三方雲端伺服器相比，機構對資料保安可掌握更多的控制權；然而，機構應判斷自身是否有足夠的專業知識安全地運作及保護內部伺服器上的系統。如機構在第三方雲端伺服器使用 AI 方案[38]，並在使用過程中處理個人資料，機構應以合約協議的方式處理以下的問題：

(i) 跨境轉移資料時（如適用）如何遵守《私隱條例》（及其他適用法律）；

38 私隱專員公署鼓勵機構參閱其有關雲端運算的資料單張以獲取更多資訊：https://www.pcpd.org.hk/tc_chi/resources_centre/publications/files/IL_cloud_c.pdf

(ii) 各方在《私隱條例》定義下作為資料使用者或資料處理者（視乎情況而定）的角色和責任；及

(iii) 各方須遵從的資料保安要求，包括實體和技術方面的控制措施。

45. 一般來說，機構應對其 AI 系統的所有組件進行全面的安全測試。實施第三方開發的 AI 方案通常需要調整機構本身的技術堆疊（tech stack），當中或會涉及保安風險。其中，開源框架（open-source frameworks）因著重透明度而公開原始碼，在機器學習中尤其常見。儘管開源軟件（open-source software）是否比閉源軟件（closed-source software）更安全仍未有定論，但有研究顯示依賴第三方開發和維護程式碼的開源機器學習框架（open-source machine learning frameworks）或會導致額外的安全風險。**無論如何，實施具開源組件的 AI 方案的機構，應在維護程式碼和管理保安風險方面遵循行業的最佳保安措施**[39]**，並留意保安建議和警報**。同樣，如機構使用應用程式介面將 AI 方案以編程方式連接至機構內部的程式及系統，應小心檢視應用程式介面的保安及遵守行業的最佳措施[40]。

確保系統安全及數據安全

46. 機構應根據所涉的風險程度，考慮採取（或在適當情況下讓 AI 供應商參與）以下措施，以確保 AI 系統穩健、可靠和安全：

(i) 實施措施（例如紅隊演練）以盡量減低機器學習模型遭受攻擊的風險，例如 AI 系統被惡意輸入 / 提示或注入訓練數據（數據中毒攻擊，data poisoning attacks），或被故意用作生成不正確或不安全的輸出結果（對抗式攻擊，adversarial attacks）[41]；

(ii) 實行員工內部指引，規定可輸入 AI 系統的內容及允許 / 禁止輸入的提示；

39 機構可參閱資訊安全網提供的資料： https://www.infosec.gov.hk/en/best-practices/business/open-source-security
40 例如，通過限制 API 可向機構的 AI 系統發出的呼叫，及進行滲透測試。
41 其他可能會影響個人資料私隱、針對 AI 模型的攻擊包括模型逆向攻擊（model inversion attacks）和成員推理攻擊（membership inference attacks）。這些攻擊或會試圖揭露訓練 / 定製數據集中包含的個人資料。

例子 3：一間律師事務所正定製第三方開發的 AI 聊天機械人，以協助其員工草擬法律文件及進行文書工作。視乎 AI 聊天機械人是在機構的內部伺服器還是雲端伺服器運行，該事務所應提醒員工在使用 AI 聊天機械人時，避免輸入個人資料及 / 或客戶的機密資訊。

(iii) 設立多重的緩衝層以防止 AI 系統不同層面或不同模組出現錯誤或故障；

(iv) 制定應變計劃（例如 AI 事故應變計劃 - 見第 3.3 節），以便有需要時能迅速暫停 AI 系統及啟動後備方案；

(v) 建立機制確保 AI 系統的運作具足夠的透明度，讓終端使用者能夠解釋系統輸出的結果；及

(vi) 建立機制讓 AI 系統輸出的結果可追溯 [42] 和可審核，例如在適當情況下和按照資料最少化原則，自動記錄 AI 系統運作時的事件（即日誌）。

圖 17：實施 AI 方案的注意事項的例子

驗證 AI 方案是否符合採購要求	測試 AI 方案
用戶接受度測試	有關託管 AI 系統的法律責任和保安考慮
防止對抗式攻擊的保安措施	確保透明度、輸出結果的可追溯性和系統可審計性的機制

42 可追溯性是指能夠記錄 AI 系統的開發及使用，包括訓練和決策過程，以及所用的數據。可追溯性通常是透過記錄存檔的形式達到。確保可追溯性有助實現可審核性。

3.3 AI 系統的管理與持續監察

主要原則：可靠、穩健及安全 / 人為監督

47. **由於使用 AI 系統的風險因素或會隨著時間而改變，機構應持續監察及檢視 AI 系統**。AI 模型本身也或會隨著時間自我學習及不斷演進，影響 AI 系統的可靠性、穩健性、真實性及安全性。

48. 高風險的 AI 系統比低風險系統需要較頻密和嚴格的監察及檢視。我們建議機構考慮採取以下的檢視機制：

(i) 對 AI 系統的採購、風險評估、所採取的緩減風險措施、數據來源、數據準備、定製、測試及驗證、實施及使用，妥善地記錄存檔，並考慮 AI 供應商是否有類似的存檔做法；

(ii) 在適當的情況下，按照資料最少化的原則監察和記錄 AI 系統的輸入內容（例如提示、查詢及要求），以助防止濫用、進行審計，以及調查任何資料外洩事故[43]；

(iii) 重新評估 AI 系統（尤其當 AI 系統的功能或運作有重大改變，或監管或科技環境出現重大變化時[44]），以識別及應對新的風險；

(iv) 定期檢視 AI 模型及在有需要和適當的情況下，考慮要求 AI 供應商檢視模型，以確保模型的運作及表現符合預期；

(v) 監察 AI 模型是否出現「模型漂移」或「模型衰退」[45]，在必要時和適當時與 AI 供應商一同糾正問題，以確保 AI 系統的輸出結果在現實世界發生變化後仍然準確（例如，定期以新數據微調及再訓練 AI 模型）；

(vi) 與 AI 供應商建立反饋及運作支援的渠道，以持續管理 AI 系統，當中可包括機構內部的 AI 系統使用者以及受 AI 系統影響的人士的意見（見第四部）；

(vii) 考慮 AI 系統的風險概況，確保對 AI 系統訂立適當程度的人為監督；

43 例如，我們建議機構根據穩健的數據管理流程處理、匿名化和適當地刪除該日誌。
44 簡單的電腦保安修補及電腦程式錯誤修正通常不會觸發重新評估 AI 系統的風險的需要。
45 「模型漂移」(model drift) 或「模型衰退」（model decay）是指，由於 AI 模型輸出結果的環境或目標變數的變化（「概念漂移」）或 AI 模型輸出結果所使用的輸入數據的變化（「數據漂移」），模型的準確性或表現隨著時間下降。

人為監督應以避免及盡量減低 AI 對個人造成的風險為目的。進行人為監督的人員應：

- 盡可能了解 AI 系統的能力和限制；
- 對過份依賴 AI 輸出結果的傾向（即「自動化的偏見」）保持警覺；
- 正確地解釋及評估 AI 輸出的結果；
- 在 AI 輸出的結果出現異常時，作出標記並在適當情況下不理會、撤銷或推翻結果；及
- 在 AI 供應商就 AI 系統輸出結果提供的資訊協助下，適時介入及中斷 AI 系統的運作。

(viii) 在 AI 系統由定製、實施、使用、監察以至終止的整個生命周期維持穩健的保安措施；及

(ix) 定期評估宏觀的科技環境，以識別機構與現時的科技生態系統的差距，並在有需要時調整 AI 策略及管治架構。

49. 我們建議機構考慮制定 **AI 事故應變計劃**，以監察和應對可能意外發生的事件[46]，計劃可包含以下要素：

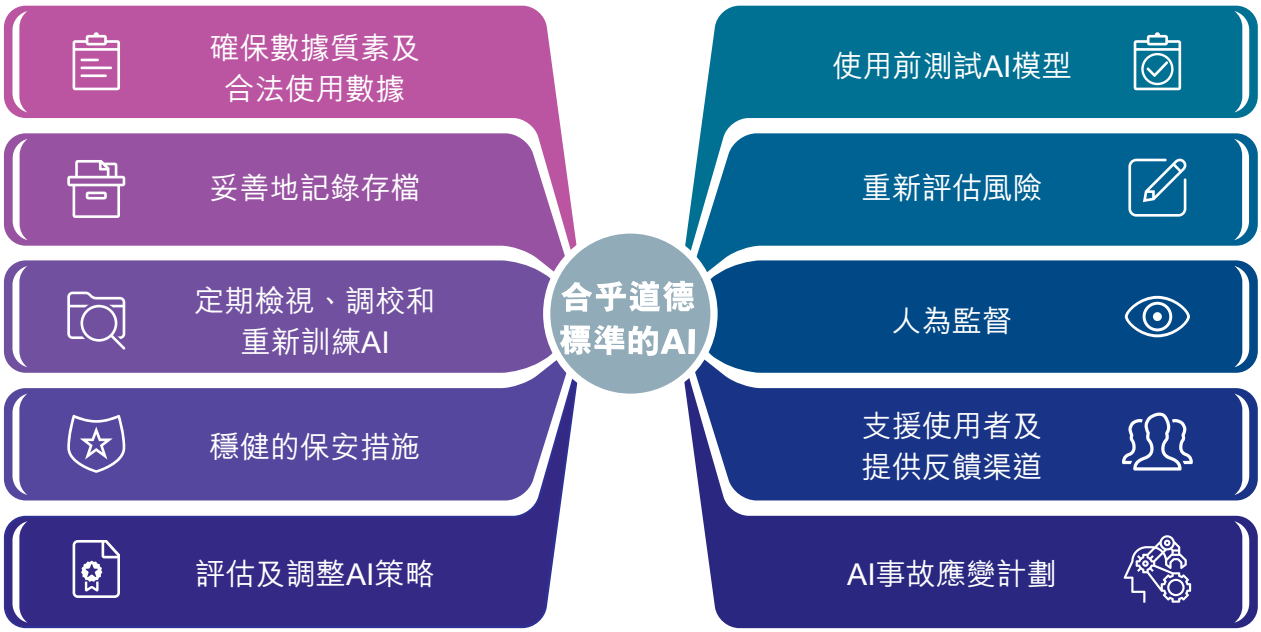
圖 18： *AI 事故應變計劃*

界定AI事故	• 機構應根據其AI系統的情況為AI事故作出定義。 • AI事故可定義為「因AI系統的開發或使用〔據稱〕對人、財產或環境造成損害的事件，包括侵犯人權（例如私隱和不受歧視的權利）；涉及人身傷害或死亡的傷害，可被視為『嚴重事件』」[47]。
監察AI事故	• 應留意和監察可預見的損害類別，並制定程序應對不可預見的損害。此步驟與風險評估過程密切相關。 • 機構可從「AI事故資料庫」得知過去發生的AI事故[48]。
通報AI事故	• 應制定內部政策和程序，以便員工就事故作出匯報，並讓其他持份者（即業務合作夥伴、客戶）透過反饋渠道通報任何事故。
遏止AI事故擴大	• 指派人員負責按既定政策和程序「暫停」或「停止」相關的AI系統，並切斷受影響的系統及其他運行的系統的連接。 • 應在切實可行的情況下盡快通知相關監管機構和受影響的人士。
調查AI事故	• 相關人員(包括負責實施AI系統的人員)應進行徹底的檢視和調查，並作出技術上的修補。 • 調查結果應根據機構的AI政策作出匯報。 • 只有在確定進一步損害或意想不到後果的風險降至最低後，才可恢復相關AI系統的運作。
從AI事故復原	• 應記錄事故調查的重點結果。 • 調查結果或會使採購AI的內部政策和程序有修訂的需要、實施和使用AI的策略有改變的需要，以及內部培訓有更新的需要。

46 如 AI 事故中涉及資料外洩事故，機構應同時進行資料外洩的應變措施。
47 經濟合作與發展組織（OECD）（2023），*《AI 事故定義的發展》*，經濟合作與發展組織 AI 文件，第 4 號。經濟合作與發展組織出版，巴黎，https://doi.org/10.1787/c323ac71-en；https://oecd.ai/en/wonk/incidents-monitor-aim
48 https://incidentdatabase.ai/

50. **機構應定期對 AI 系統進行內部審核（及有需要時進行獨立評估），以確保 AI 的使用持續遵從機構相關政策的規定，以及與 AI 策略保持一致。**審核結果應向機構的董事會、高級管理層及管治組織（例如審核委員會）匯報。

圖 19： ***AI 系統管理***



第四部 與持份者的溝通及交流

主要原則：透明度與可解釋性

4.1 提供資訊

51. 機構應清楚讓持份者知道機構在使用 AI，以展示機構奉行「透明度與可解釋性」原則。**機構應定期及有效地與持份者（尤其是內部員工、AI 供應商、個別消費者及監管機構）聯絡及交流。有效的溝通對於建立信任至關重要**。

52. 如定製及使用 AI 的過程涉及個人資料，機構必須依據《私隱條例》的保障資料第 1(3) 及第 5 原則，向相關資料當事人告知所需的資訊，當中包括但不限於：

(i) 個人資料將會用於甚麼目的，例如用於訓練及 / 或定製 AI 模型、促進自動化決策等；
(ii) 資料可能會轉移予甚麼類別的人士，例如 AI 供應商；及
(iii) 機構在定製及使用 AI 方面的個人資料政策及行事方式。

53. 此外，為提高透明度與公開性，在與持份者溝通時（特別是員工、個別消費者及監管機構），機構應考慮採取以下步驟：

(i) 除非在有關情況和背景下 AI 系統的使用是顯而易見的，否則，機構應清楚地及用顯著的方式披露 AI 系統的使用；
(ii) 機構應提供足夠的資訊[49]，說明在其產品或服務中使用 AI 系統的目的、益處、限制及效果[50]；及
(iii) 機構應披露 AI 系統的風險評估結果[51]。

49 機構可考慮透過 AI 模型卡（AI model card）披露 AI 系統的相關資訊。AI 模型卡是「機器學習模型提供的簡短文檔，解釋模型的使用背景、表現評估程序的細節和其他相關資訊。」（來源：IAPP（原文為英文））
50 除非有關披露會對商業敏感或專有資訊造成損害。
51 除非有關披露會對商業敏感或專有資訊造成損害。

54. 若 AI 供應商較機構更為合適提供上述資訊（特別是有關 AI 系統的技術資訊），機構可在整個採購過程及之後與 AI 供應商密切協調，並在必要時要求他們提供相關指引或尋求他們的專業知識來應對持份者的關注。

4.2 資料當事人的權利及反饋

55. 如使用 AI 的機構涉及處理個人資料，機構必須注意資料當事人有權分別根據《私隱條例》第 18 和第 22 條提出查閱資料要求和改正資料要求。機構可在必要時與 AI 供應商交流以滿足這些要求。

56. 如 AI 系統的決策 / 輸出結果可能對個人造成重大影響，機構應盡可能向個人提供途徑，讓他們作出反饋、尋求解釋及 / 或要求人為介入。機構亦應仔細考慮是否向個人提供退出使用 AI 系統的選項。

57. 從更廣泛的角度來看，我們建議機構設立供內部員工及 / 或個別消費者表達意見的渠道，並鼓勵作出反饋，以就相關 AI 系統作出調整及 / 或將反饋傳達給 AI 供應商（如適用）。

4.3 可解釋的 AI

58. 讓 AI 的決策及輸出結果達致可解釋性是建立持份者信任的關鍵。在可行的情況下，機構提供的解釋可包括下列資訊（尤其當 AI 系統的使用可能對個人造成重大影響時）[52]：

(i) AI 如何參與決策過程及其參與的程度，包括所採用的 AI 系統主要負責的工作的概述，以及人類決策者的參與情況（如有）；

(ii) 個人資料在自動化或 AI 輔助的決策過程或內容生成過程中如何被使用，以及為何該資料被視為相關和必需；及

(iii) 促使 AI 系統作出自動化決策 / 輸出結果的主要因素（全局可解釋性），以及導致個別決策 / 輸出個別結果的主要因素（局部可解釋性）。如提供解釋是不可行的，應明確說明。

52 有關如何有意義地解釋 AI 所作的自動化決策，機構可考慮參考英國資訊專員辦公室與阿蘭 · 圖靈研究所於 2020 年發出的《*解釋 AI 決策*》指引，以取得更多建議：https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/explaining-decisions-made-with-artificial-intelligence/。

圖 20： ***與持份者的溝通及交流***

59. 若 AI 供應商較機構更適合就 AI 系統的決策及所輸出的結果作出解釋，機構可考慮在適當的情況下與 AI 供應商合作。機構在決定披露的資訊類別及詳細程度時，應考慮有關持份者理解相關資訊的能力、他們的需要，以及有關披露會否對 AI 系統的保安和合法目的造成不利影響等。例如，就用作偵測顧客欺詐行為及其他犯罪行為的 AI 系統而言，有關機構或毋須披露 AI 系統的相關指標，以免顧客有機會知悉如何繞過該系統。另一方面，若 AI 系統用作協助員工內部使用並根據內部數據進行定製，機構可在可行及適當的情況下考慮提供選項，容許員工追蹤 AI 系統輸出的結果 / 決策的資訊來源，以確保其準確性 [53]。

4.4 語言及方式

60. 與持份者（尤其消費者）溝通，應使用淺白的語言和清楚易明的方式，並設法讓持份者知悉。有關資訊亦可納入機構的私隱政策內。

與持份者溝通，應使用淺白的語言清楚易明的方式表達，並設法讓持份者知悉。

53 例如，如定製過程涉及使用檢索增強生成技術。

鳴謝

個人資料私隱專員公署（私隱專員公署）衷心感謝以下個人和組織以及主要AI 供應商在諮詢過程中提供寶貴意見（按英文名稱排序）：

支持機構

香港應用科技研究院

政府資訊科技總監辦公室

私隱專員公署科技發展常務委員會成員

陳仲文工程師，香港生產力促進局數碼轉型部總經理

張偉倫先生，香港應用科技研究院人工智能及可信技術部門首席總監

劉偉經特邀教授，國際信息系統審計會（ISACA）全球董事會成員

李嘉樂博士，創星滙（香港）創會主席兼總裁

黃錦輝議員，MH，香港中文大學工程學院副院長（外務）

姚兆明教授，香港大學計算機科學系教授及副系主任

機構

香港大學 AI & Humanity Lab

亞洲證券業與金融市場協會

資訊政策領導中心

德勤

安永

香港銀行公會

香港電腦學會

香港金融管理局

香港生產力促進局

香港科技園公司

香港城市大學香港持續發展研究中心

附錄 A - 《個人資料（私隱）條例》的保障資料原則

《個人資料（私隱）條例》（第 486 章）（《私隱條例》）規管公私營機構收集、持有、處理及使用個人資料的情況。《私隱條例》屬於科技中立及原則性的法例。《私隱條例》附表 1 的保障資料原則是《私隱條例》的核心規定，涵蓋由收集至銷毀整個處理個人資料的生命周期。

保障資料第 1 原則 — 收集目的及方式

保障資料第 1 原則訂明，資料使用者只可為直接與其職能或活動有關的合法目的，收集個人資料；而收集的方法須是合法和公平的；收集的資料對該目的而言須要屬必需及足夠的，但不可超乎適度。

資料使用者亦須清楚表明收集資料的目的、資料可能會被轉移給哪類人士，以及資料當事人可要求查閱和改正自己的資料的權利及途徑。有關資訊通常在《收集個人資料聲明》中呈列。

保障資料第 2 原則 — 準確性及保留期間

保障資料第 2 原則要求資料使用者採取所有切實可行的步驟，以確保持有的個人資料準確無誤，而保留時間不超過將其保存以貫徹該資料被使用於的目的所需的時間。《私隱條例》第 26 條亦有類似規定，要求資料使用者刪除不再需要的個人資料。

如資料使用者聘請資料處理者處理個人資料，資料使用者須採取合約規範方法或其他方法，以防止資料處理者將個人資料保存超過所需的時間。

保障資料第 3 原則 — 資料的使用

保障資料第 3 原則訂明，除非得到資料當事人自願給予的明示同意，否則個人資料不得用於新目的，即與原本的收集目的不同或不相關的目的。

保障資料第 4 原則 — 資料的保安

保障資料第 4 原則要求資料使用者採取所有切實可行的步驟，保障其持有的個人資料不會未經授權或意外地被查閱、處理、刪除、喪失或使用。

如資料使用者聘用資料處理者處理個人資料，必須採取合約規範方法或其他方法，確保資料處理者依從上述的資料保安要求。

保障資料第 5 原則 — 透明度

保障資料第 5 原則訂明，資料使用者須採取所有切實可行的步驟，確保某些資訊在一般情況下可提供予公眾，包括在個人資料方面的政策及實務，所持有的個人資料的種類和主要使用於甚麼目的。

保障資料第 6 原則 — 查閱及改正

保障資料第 6 原則賦予資料當事人可要求查閱及改正自己的個人資料的權利。

《私隱條例》的第 5 部另有詳細條文補充保障資料第 6 原則的規定，具體訂明遵從查閱及改正資料要求的方式及時限，以及在甚麼情況下資料使用者可拒絕依從這些要求等。

附錄 B - 主要參考資料

- 新加坡資訊通信媒體發展局與 Aicadium，*《生成式人工智能：信任與治理的影響》*(2024)[54]
- 新加坡資訊通信媒體發展局與人工智能驗證基金會，*《針對生成式人工智能的人工智能管理建議模範框架促進受信賴的生態系統》*(2024)[55]
- 國際信息系統審計協會，*《人工智能：機器學習、深度學習與神經網路概論》*(2024)[56]
- 經濟合作與發展組織，*《人工智能原則》*(2024)[57]
- 新加坡個人資料保護委員會，*《關於在人工智能推薦與決策系統中使用個人資料的指引》*(2024)[58]
- 法國資料保障機構，*《人工智能操作指南》*(2023)[59]
- 英國競爭及市場管理局，*《人工智能基礎模型：初期報告》*(2023)[60]
- 中華人民共和國國家互聯網信息辦公室，*《全球人工智能管治倡議》*(2023)[61]
- 英國資訊專員辦公室，*《人工智能及資料保障指引》*(2023)[62]
- 國際信息系統審計協會，*《人工智能革命的前景與風險：管理風險》*(2023)[63]
- 國際標準化組織，*《ISO/IEC 23894:2023 資訊科技 - 人工智能 - 風險管理指南》*(2023)[64]
- 國際標準化組織，*《ISO/IEC 42001:2023 資訊科技 - 人工智能管理系統》*(2023)[65]
- Meta，*《Llama 2 - 負責任使用指南》*(2023)[66]

54 https://aiverifyfoundation.sg/downloads/Discussion_Paper.pdf
55 https://aiverifyfoundation.sg/downloads/Proposed_MGF_Gen_AI_2024.pdf
56 https://store.isaca.org/s/store#/store/browse/detail/a2S4w000008Kn59EAC
57 https://oecd.ai/en/ai-principles
58 https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/advisory-guidelines/advisory-guidelines-on-the-use-of-personal-data-in-ai-recommendation-and-decision-systems.pdf
59 https://www.cnil.fr/en/ai-how-sheets
60 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1185508/Full_report_.pdf
61 https://www.mfa.gov.cn/eng/wjdt_665385/2649_665393/202310/t20231020_11164834.html
62 https://ico.org.uk/media/for-organisations/guide-to-data-protection/key-dp-themes/guidance-on-ai-and-data-protection-2-0.pdf
63 https://www.isaca.org/resources/white-papers/2023/the-promise-and-peril-of-the-ai-revolution
64 https://www.iso.org/obp/ui/en/#iso:std:iso-iec:23894:ed-1:v1:en
65 https://www.iso.org/standard/81230.html
66 https://llama.meta.com/responsible-use-guide/

- 英國國家網絡安全中心與美國網絡安全和基礎設施安全局，*《開發安全人工智能系統指南》*(2023)[67]
- 美國國家標準暨技術研究院，*《人工智能風險管理框架 (AI RMF 1.0)》*(2023)[68]
- 中華人民共和國全國信息安全標準化技術委員會秘書處，*《網絡安全標準實踐指南 - 生成式人工智能服務內容標識方法》*(2023)[69]
- 經濟合作與發展組織，*《推進人工智能的問責：可信賴的人工智能於整個生命週期的風險管治管理》*(2023)[70]
- 中國香港特別行政區政府資訊科技總監辦公室，*《人工智能道德框架》*(供公眾參考特製版本)(2023 修訂版)[71]
- 加拿大私隱專員公署，*《負責任、可信賴和保護私隱的生成式人工智能技術原則》*(2023)[72]
- 聯合國人工智能諮詢機構，*《臨時報告：為人類治理人工智能》*(2023)[73]
- 聯合國教科文組織，*《人工智能倫理建議書》*(2023)[74]
- 世界經濟論壇，*《負責任地採用人工智能：私營企業採購人工智能解決方案指南》*(2023)[75]
- 英國資訊專員辦公室，*《人工智能與資料保障風險實務手冊》*(2022)[76]
- 國際資訊系統審計協會，*《制定人工智能管治框架》*(2022)[77]
- 微軟，*《微軟負責任的 AI 標準 (第二版) 一般要求》*(2022)[78]
- Wiley，*《值得信賴的 AI - 在人工智能中探索信任與道德的商業指南》*(2022)[79]

67 https://www.ncsc.gov.uk/files/Guidelines-for-secure-AI-system-development.pdf
68 https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf
69 https://www.tc260.org.cn/upload/2023-08-25/1692961404507050376.pdf
70 https://www.oecd.org/sti/advancing-accountability-in-ai-2448f04b-en.htm
71 https://www.ogcio.gov.hk/en/our_work/infrastructure/methodology/ethical_ai_framework/doc/Ethical_AI_Framework.pdf
72 https://www.priv.gc.ca/en/privacy-topics/technology/artificial-intelligence/gd_principles_ai/
73 https://www.un.org/en/ai-advisory-body
74 https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence
75 https://www.weforum.org/publications/adopting-ai-responsibly-guidelines-for-procurement-of-ai-solutions-by-the-private-sector/
76 https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/ai-and-data-protection-risk-toolkit/
77 https://www.isaca.org/resources/news-and-trends/newsletters/atisaca/2022/volume-38/developing-an-artificial-intelligence-governance-framework
78 https://blogs.microsoft.com/wp-content/uploads/prod/sites/5/2022/06/Microsoft-Responsible-AI-Standard-v2-General-Requirements-3.pdf
79 https://www.wiley.com/en-us/Trustworthy+AI%3A+A+Business+Guide+for+Navigating+Trust+and+Ethics+in+AI-p-9781119867951

附錄 B - 主要參考資料

- 中華人民共和國國家新一代人工智能治理專業委員會，*《新一代人工智能倫理規範》*（2021）[80]
- 英國人工智能辦公室，*《人工智能採購指南》*（2021）[81]
- 中國香港特別行政區個人資料私隱專員公署，*《開發及使用人工智能道德標準指引》*（2021）[82]
- 新加坡資訊通信媒體發展局與新加坡個人資料保護委員會，*《人工智能管理模範框架》*（2020 第二版）[83]
- 英國資訊專員辦公室與阿蘭 · 圖靈研究所，*《解釋人工智能決策》*（2020）[84]
- 谷歌，*《負責任的人工智能實務指引》* [85]
- 國際私隱專業人員協會，*《人工智能理關鍵術語》* [86]

80 https://www.most.gov.cn/kjbgz/202109/t20210926_177063.html
81 https://www.gov.uk/government/publications/guidelines-for-ai-procurement/guidelines-for-ai-procurement
82 https://www.pcpd.org.hk/english/resources_centre/publications/files/guidance_ethical_e.pdf
83 https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/SGModelAIGovFramework2.pdf
84 https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/explaining-decisions-made-with-artificial-intelligence/
85 https://ai.google/responsibility/responsible-ai-practices/
86 https://iapp.org/resources/article/key-terms-for-ai-governance/

香港個人資料私隱專員公署
Office of the Privacy Commissioner for Personal Data, Hong Kong

電話：2827 2827
傳真：2877 7026
地址：香港灣仔皇后大道東248號大新金融中心13樓1303室
電郵：communications@pcpd.org.hk

私隱專員公署網頁
pcpd.org.hk

下載本刊物

免責聲明

本刊物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，亦不構成法律或其他專業意見。私隱專員並沒有就本刊物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。

二零二四年六月