

# 開發及使用 人工智能道德標準指引



PCPD



HK



[PCPD.org.hk](http://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



守護 · 私隱 · 廿五載  
GUARDIAN · PRIVACY · 25 YEARS



# 目錄

<b>1 引言</b>	<b>2</b>
<b>2 數據管理價值</b>	<b>7</b>
2.1 尊重	7
2.2 互惠	7
2.3 公平	7
<b>3 人工智能的道德原則</b>	<b>8</b>
3.1 問責	8
3.2 人為監督	8
3.3 透明度與可解釋性	8
3.4 數據私隱	9
3.5 公平	9
3.6 有益的人工智能	9
3.7 可靠、穩健及安全	9
<b>4 實務指引</b>	<b>11</b>
4.1 人工智能策略及管治	12
4.1.1 人工智能策略	12
4.1.2 管治架構	13
4.1.3 培訓及加強認識	15
4.2 風險評估及人為監督	16
4.2.1 須考慮的風險因素	18
4.2.2 決定人為監督的程度	20
4.3 人工智能模型的開發及人工智能系統的管理	22
4.3.1 為人工智能準備數據	23
4.3.2 人工智能模型的開發	26
4.3.3 人工智能系統的管理與監察	27
4.4 與持份者的溝通及交流	30
<b>5 由第三方提供的人工智能系統</b>	<b>32</b>
<b>6 結語</b>	<b>33</b>
<b>附錄 A - 自我評估核對清單</b>	<b>34</b>
<b>附錄 B - 《個人資料(私隱)條例》的保障資料原則</b>	<b>38</b>
<b>附錄 C - 參考資料</b>	<b>40</b>

# 1 引言



## 何謂人工智能？

人工智能指一系列涉及以電腦程式和機器模仿人類解難及決策能力的科技。人工智能的應用例子包括臉容識別、語音識別、聊天機械人、數據分析、自動化決策或建議等。人工智能科技仍在演變中，更多新的應用或會出現。

## 本指引的適用範圍和目的

開發及使用人工智能時常會用到個人資料。當開發或使用人工智能系統時涉及使用個人資料，或涉及識別、評估或監察個人，這份《開發及使用人工智能道德標準指引》(指引)便適用，皆因兩者都可能影響到個人資料私隱。

本指引的目的是促進在香港健康發展和使用人工智能，並協助機構在開發及使用人工智能時遵從《個人資料(私隱)條例》(第486章)(《私隱條例》)的規定。

在本指引中，「人工智能」一詞泛指有關科技；「人工智能模型」指利用數據集建立及訓練的數學演算法；而「人工智能系統」指機構用來協助營運的電腦程式。雖然這幾個詞語的實際意思有別，但在本指引中的意思可相通。本指引所建議的價值、原則及措施在實質上不會因使用有關詞語而受到影響。

本指引的附錄A是一份自我評估核對清單，協助機構判斷它們在開發及使用人工智能時是否已採納本指引所建議的措施。

## 人工智能的好處

越來越多機構（包括商業實體、政府部門及公營機構）在營運時使用人工智能。銀行利用人工智能評估客戶的信用可靠程度及偵測洗黑錢活動。醫療服務提供者利用人工智能分析醫療紀錄及協助醫生診斷。政府部門利用人工智能監察及完善道路交通，以減少擠塞情況。其他機構亦利用人工智能評估求職者的履歷、回覆顧客的查詢等。人工智能透過節省人手、改善運作效率、優化資源分配、提供個人化服務以及引發新構想，為不同行業帶來龐大機遇和利益。有研究顯示，到2030年人工智能將為全球的本地生產總值提升14%<sup>1</sup>。

## 人工智能的風險

人工智能的潛能是透過這個數碼年代不斷產生的大數據而實現，因此在開發及使用人工智能的過程中時常涉及個人資料。尤其是新一代的人工智能，其「智能」是透過利用複雜的機器學習演算法分析大量訓練數據而獲得的。因此，人工智能不斷在測試通用的保障資料原則（例如透明度、數據最少化及使用限制）的規限，對私隱及個人資料保障帶來挑戰。此外，人工智能的資料保障風險與其潛在的道德方面的影響有相通之處，因為當個人的個人資料被人工智能系統分析，他們的權利、自由及利益亦可能因人工智能系統所作的自動化決策而受到影響。因此，如人工智能被不當使用，或會對人權（包括私隱權）、人類尊嚴、個人自主及公平造成損害。使用人工智能的機構可能因此而失去消費者的信任。

## 遵從《個人資料（私隱）條例》

個人資料屬於個人，其收集、持有、處理及使用受到《私隱條例》所規管。機構須注意，它們在開發及使用人工智能的過程中須根據《私隱條例》合法地收集、持有、處理及使用個人資料。本指引的附錄B概述《私隱條例》附表1的六項保障資料原則。該六項保障資料原則代表《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。

## 開發及使用人工智能的道德標準

鑑於人工智能潛在的道德方面的風險，機構亦應在其營運以及開發及使用人工智能的過程中，秉持良好的數據道德標準。因此，機構應顧及所有相關持份者的權利、自由及利益（即採取多方持份者參與的方式），以減低私隱風險及道德方面的風險。

<sup>1</sup> 普華永道，衡量人工智能的機遇 — 人工智能對業務的真正價值為何和如何把握機遇 (Sizing the Prize - What's the real value of AI for your business and how can you capitalise?) (2017)

有見及此，近年要求負責任及有道德地使用人工智能的呼聲有所增加。世界各地亦紛紛推出有關使用人工智能的原則和指引。例如，環球私隱議會<sup>2</sup>、歐盟委員會<sup>3</sup>、經濟合作與發展組織<sup>4</sup>、聯合國教科文組織<sup>5</sup>、日本<sup>6</sup>及新加坡<sup>7</sup>在近幾年都各自發出相關指引。這些指引中可以見到一些共同原則，例如問責性、透明度、公平、數據私隱及人為監督，顯示全球在這方面的共識。歐盟委員會於2021年4月就立法規管人工智能提出法案<sup>8</sup>。如法案獲通過，有機會成為世界上首條規管人工智能的法例。

圖 1 近期環球人工智能管治發展的時間線



2 環球私隱議會為全球超過 130 個資料保障機構提供一個領先的國際平台，就私隱議題和國際最新發展進行討論和交流。環球私隱議會於 2018 年採納了《人工智能的倫理道德與資料保障宣言》，支持利用六項指導原則促使在開發人工智能的過程中能保護人類的權利。在 2020 年，環球私隱議會採納了《開發及應用人工智能的體現問責決議》，建議開發及使用人工智能的機構採取 12 項問責措施，以期與持份者建立信任。

3 歐盟委員會-人工智能獨立高級專家組，《可信賴的人工智能的道德準則》(2019)

4 經濟合作與發展組織，《經合組織理事會有關人工智能的建議》(2019)

5 聯合國教科文組織，《人工智能道德建議書草案文本初稿》(2020)

6 日本，《以人為本的人工智能社會原則》(2019)

7 新加坡，《人工智能管理模範框架》(第 1 版) (2019)。第 2 版於 2020 年出版。

8 歐盟委員會，《關於制定人工智能統一規則的條例提案》(2021)

本指引是以上述有關使用人工智能的原則及指引作為基礎，並包含香港個人資料私隱專員公署（私隱公署）自2019年以來擔任環球私隱議會的人工智能的道德與數據保障工作小組的聯席主席所獲得的經驗。本指引建議一套數據管理價值及人工智能的道德原則，亦提供一套按照一般業務程序而撰寫的實務指引，協助機構以合法（就《私隱條例》而言）和合乎道德標準的方式開發及使用人工智能，讓機構取得持份者（尤其是個別的消費者）的信任。

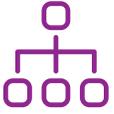
隨著歐盟委員會最近作出立法提案，我們留意到對於人工智能應否以法律或其他方式規管以及規管的程度如何，在社會上尚未取得共識。在英國，甚至有聲音提出為了促進創新和發展人工智能，應取消該國保障資料法例下容許個人拒絕受制於完全自動化決策的條文。在香港，我們相信就開發及使用人工智能提供私隱友善和合乎道德標準的行事指引，可促進創新及擴闊人工智能在社會上的應用。

香港正致力成為大灣區及亞太區的數據樞紐和創新中心，以及世界級的智慧城市。由於數據是人工智能的命脈，香港可利用及發揮其作為數據樞紐的優勢以推動人工智能的發展。健康地發展和使用人工智能亦可大大促進香港成為創新中心和世界級的智慧城市。我們相信本指引能協助香港的機構在開發及使用人工智能方面開啟成功之門。

圖 2 本指引的架構



# 2 數據管理價值



機構的價值觀會影響有關機構如何進行活動以達致其使命和願景。機構要確保開發及使用人工智能合乎道德標準，首先應界定其核心道德價值。私隱公署於2018年10月出版的《中國香港的道德問責框架》<sup>9</sup>建議機構採納三項數據管理價值，分別是尊重、互惠及公平。這三項價值是制定有關開發及使用人工智能的道德原則和措施的出發點。

## 2.1 尊重

在處理個人資料時，尊重有關個人的尊嚴、自主、權利、利益以及合理期望是至關重要。因此，每名個人都應得到有道德的對待，而不應被視為一個物件或一項數據。

## 2.2 互惠

「互惠」強調要盡可能向持份者（包括受使用人工智能影響的個人及社會大眾）帶來益處，同時應避免或減低對持份者造成傷害。

## 2.3 公平

「公平」應同時體現於過程和結果當中。在過程方面，「公平」是指合理地作出決定，沒有不義的偏見或非法的歧視。機構應建立容易採用及有效的途徑，讓受到不公平對待的人士尋求糾正或補償。在結果方面，「公平」指相類似的人士應獲得相類似的對待。若不同個人或不同群組之間的待遇有任何差異，應該要有充份的理據支持。

<sup>9</sup> 見：[https://www.pcpd.org.hk/misc/files/Ethical\\_Accountability\\_Framework.pdf](https://www.pcpd.org.hk/misc/files/Ethical_Accountability_Framework.pdf) (只有英文版)

# 3 人工智能的道德原則



在考慮了上述三項數據管理價值及其自身的企業價值觀後，機構應制定相容的原則及政策以彰顯有關價值。因此，我們鼓勵機構採納以下人工智能的道德原則。

## 3.1 問責

機構應對其作為負責，並且能夠就其行為提供充分的理據。機構應制定措施，以評估及應對人工智能的風險，過程中應有高級管理層的參與及跨專業領域的合作。

## 3.2 人為監督

人工智能系統的使用者應能就人工智能系統的建議或決定作出知情及自主的行動。在過程中，人為參與的程度應與使用人工智能系統的風險及影響相稱。如使用人工智能被評估為高風險，應該確保可以有人為介入。

## 3.3 透明度與可解釋性

機構應清楚明確地披露它們使用人工智能以及相關的數據私隱措施，並致力改善自動化和人工智能輔助作出的決定的可解釋性<sup>10</sup>。透明度與可解釋性有助彰顯問責性，以及保障個人在使用人工智能時的權利、自由及利益。

<sup>10</sup> 可解釋性是指從人工智能系統中找出因果關係的能力。換言之，它是指當人工智能系統的輸入數據或演算法參數有所更改時，一個人有多大程度能預測到會出現的新境況。

### 3.4 數據私隱

私隱是基本人權。機構應具備有效的數據管治，務求在開發及使用人工智能時保障個人的私隱。處理及保障在開發及使用人工智能所涉及的個人資料時，應依從《私隱條例》的規定，尤其是附表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，涵蓋由收集、保留、使用以至刪除的整個處理個人資料的生命周期。保障資料原則的詳情見附錄B。

### 3.5 公平

個人有權獲得合理地相等的方式對待，不應遭受不義的偏見或非法的歧視。若不同個人或不同群組之間的待遇有任何差異，應該要有充份的理據支持。

### 3.6 有益的人工智能

人工智能應該為人類、企業及社會大眾帶來益處，當中包括防止傷害。如使用人工智能可能對持份者造成傷害，機構應採取措施減低造成傷害的可能性及嚴重性。

### 3.7 可靠、穩健及安全

機構應確保人工智能系統在預計的系統使用期限內可靠地運作。人工智能系統應能在運作過程中應付錯誤，以避免或減低因意外而造成的傷害。人工智能系統亦應得到保護而免遭攻擊，例如黑客入侵及數據下毒<sup>11</sup>。機構應制定應變計劃，以為人工智能系統不能正常運作時作好準備。

11 數據下毒 (data poisoning) 是對人工智能系統的一種攻擊，當中透過污染訓練數據而影響系統作出正確預測的能力。

**圖 3 數據管理價值及人工智能的道德原則的配對**

	數據管理價值	人工智能的道德原則
1	尊重	<ul style="list-style-type: none"><li>• 問責</li><li>• 人為監督</li><li>• 透明度與可解釋性</li><li>• 數據私隱</li></ul>
2	互惠	<ul style="list-style-type: none"><li>• 有益的人工智能</li><li>• 可靠、穩健及安全</li></ul>
3	公平	<ul style="list-style-type: none"><li>• 公平</li></ul>

# 4 實務指引



為確保上述的價值及道德原則可以在運作中落實執行，機構應制定適當的政策、措施及程序。因此，機構在開發及使用人工智能的整個業務流程中應考慮下述範疇的建議措施：

- 人工智能策略及管治；
- 風險評估及人為監督；
- 人工智能模型的開發及人工智能系統的管理；及
- 與持份者的溝通及交流。

圖 4 合乎道德標準地開發及使用人工智能的工作流程



一般來說，機構應以風險為本的方式開發、管理及使用人工智能系統。因此在考慮及採用下述建議措施時，機構應務求所採取的措施與人工智能系統可能構成的風險相稱。下述建議並非詳盡無遺，機構在開發及使用人工智能時應採取其他適當措施，以落實數據管理價值及人工智能的道德原則。

## 4.1 人工智能策略及管治

高級管理層的支持和積極參與是實施人工智能系統的成功要素。因此，機構應建立內部管治架構，以引領人工智能的開發及使用。內部管治架構一般包含機構層面的人工智能策略和人工智能管治委員會（或類似組織）。

由於人工智能系統的開發及使用需要大量數據（通常包括個人資料），機構應制定政策，務求在人工智能的生命周期中落實貫徹私隱及數據保障的設計。機構可考慮利用及修改現有關於處理個人資料的數據管治或問責框架，例如私隱公署倡議的私隱管理系統，並把本指引內的要素納入現有的工作流程，以便更容易地管理人工智能系統的開發及使用。

機構應制定人工智能策略及成立人工智能管治委員會（或類似組織），以引領人工智能的開發及使用。

### 4.1.1 人工智能策略

#### 主要原則：問責

機構應制定人工智能策略，以展示高級管理層有決心通過合乎道德標準的方式開發及使用人工智能。人工智能策略亦應包含使用人工智能的目的以及如何使用人工智能的相關指引。

機構的人工智能策略可包含下述要素：

- (i) 訂明使用人工智能系統的業務目標，例如計劃利用人工智能系統幫助解決哪些問題；
- (ii) 界定人工智能系統在機構的科技生態系統中提供的功能；
- (iii) 參考上述介紹的人工智能的道德原則，制定適用於機構在開發及使用人工智能方面的道德原則；
- (iv) 列明人工智能系統的可接受用途，以及哪些用途是不容許的。機構可就人工智能的應用採取交通燈機制<sup>12</sup>；
- (v) 確保人工智能系統的使用符合機構的願景、使命及價值；
- (vi) 就如何合乎道德標準地設計、開發及使用人工智能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；及
- (vii) 定期把人工智能策略、政策和程序傳達予所有相關人士，包括內部各級職員和外部持份者（如適當），例如業務夥伴。

#### 4.1.2 管治架構

##### 主要原則：問責 / 人為監督

開發及使用人工智能需要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，以引領人工智能策略的實施，並監督人工智能的開發及使用。人工智能的管治架構可包含下述要素：

- (i) 人工智能管治委員會（或類似組織）監督人工智能由開發、使用以至終止的整個生命周期；

<sup>12</sup> 在交通燈機制下，人工智能的實際應用會分為紅、黃、綠三個類別。紅色類別是被禁止採用的人工智能，因其所涉及的風險太高。綠色類別是低風險的人工智能，這類別的人工智能在採用前無須進行嚴格的風險評估。黃色類別包含沒有在紅色或綠色類別出現的所有其他情況。要決定黃色類別內的人工智能應否獲准使用，應先進行嚴格的風險評估。

### 人工智能管治委員會

高級管理層參與以及跨專業領域合作應是人工智能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、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、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、私隱總監或類似職位）領導該跨部門團隊。

（選擇性措施）人工智能管治委員會可向外部專家尋求人工智能及道德標準方面的獨立意見。如果某項目規模龐大、影響廣泛、備受注目，其道德價值有可能受到挑戰，機構亦可成立另一個人工智能道德委員會以進行獨立檢視。

- (ii) 就人工智能的開發及使用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：

- 系統分析師、系統架構師及數據科學家應聚焦於人工智能系統的設計、開發、監察及維護；
- 法律及合規專業人員應聚焦於確保機構在人工智能的開發及使用上遵從法律及規例（包括資料保障法例）的要求，以及內部政策的規定；
- 業務運作人員應按機構的政策和程序使用人工智能；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開發及使用人工智能；及

**需要足夠資源的例子：**

- 聘請具備相關技能、經驗及專門知識的內部和外部專家以開發及使用人工智能系統；
- 在有需要時進行風險評估，以識別及減低使用人工智能所帶來的風險，包括私隱及保安風險；
- 建立能夠協助機構監察、記錄及檢視人工智能系統的資訊系統；及
- 為相關人員提供足夠的培訓（見下文第4.1.3段）。

(iv) 就開發及使用人工智能建立有效的內部通報機制，例如通報系統故障或提出有關資料保障或道德標準上的關注，以便人工智能管治委員會能恰當地作出監察。

### 4.1.3 培訓及加強認識

**主要原則：問責**

良好的策略、計劃或政策要成功，有賴能幹的人員的執行。機構要妥善地執行有關人工智能的策略及政策，就應為所有相關人員提供相關和足夠的培訓，以確保他們具有適當的知識、技能和認識，以便在使用人工智能系統的環境中工作。培訓的例子包括：

- (i) 為系統分析師、系統架構師及數據科學家提供有關遵從法律、規例、內部政策，以及網絡保安風險方面的培訓；
- (ii) 為法律及合規專業人員和人工智能使用者（包括業務運作人員）提供人工智能科技的培訓；及
- (iii) 為負責監督人工智能系統決策的審查員提供培訓，讓他們能夠在人工智能系統所作的決定中查找並糾正不義的偏見、非法的歧視和錯誤。

機構要確保審查員認真履行其職責，以及人為監督並非只是擺姿態的性質，相關人員應有能力權衡及理解人工智能所作的建議。審查員亦應能恰當地行使酌情權和權力，在有需要時否決人工智能所作的建議。

除了以上的核心人員外，推行人工智能系統亦會涉及機構的其他部門。機構亦應提升所有相關人員對機構的人工智能策略和政策，以及人工智能風險的認識。有關方法的例子：

- (i) 向工作上與人工智能系統有關但無須直接與人工智能系統互動的人員（例如客戶服務及公關人員）提供一般簡介或培訓，讓他們了解機構所使用的人工智能系統的好處、風險、功能和限制；及
- (ii) 透過職員會議或其他內部溝通方式（例如通告）向所有相關的人員傳達合乎道德標準的人工智能和相關應用原則的重要性，以及在開發及使用人工智能方面培養和推廣尊重及有道德的文化。

## 4.2 風險評估及人為監督

不同的人工智能系統有不同的風險程度，影響風險高低的因素包括使用人工智能系統的目的及如何使用人工智能系統。例如，用來評估個人的信用可靠程度的人工智能系統一般比用於個人化廣告的人工智能系統有較高風險，因為前者或會令個人無法得到信貸安排，而後者未必會對個人造成重大影響。此外，完全自主的人工智能系統亦可能比只向人類行動者提供建議的人工智能系統有較高風險。因此，在管理人工智能系統方面，應採取風險為本的方式。在這方面，機構需要就開發及使用人工智能進行全面的風險評估，有系統地識別、分析及評估風險，包括私隱風險。對於高風險的人工智能系統，機構應在整個人工智能的生命周期內建立及持續實施風險管理機制，並將相關資料記錄存檔。

機構需要就開發及使用人工智能進行全面的風險評估，有系統地識別、分析及評估風險，包括私隱風險。

在開發及使用新的人工智能系統前，或對現時的人工智能系統進行重大更新前，應由以不同部門人員組成的跨部門團隊進行風險評估。視乎有關情況，風險評估可能需要涉及不同社會、文化和宗教背景，以及不同性別及種族的人士，以便在開發人工智能的過程中識別不義的偏見和非法的歧視。所有風險評估應妥為記錄存檔，而風險評估的結果應由人工智能管治委員會（或類似組織）檢視及認可。

圖 5 風險評估的程序



## 4.2.1 須考慮的風險因素

主要原則：有益的人工智能 / 數據私隱 / 公平

由於開發及使用人工智能時常涉及使用個人資料，因此必須應對資料私隱風險。從保障個人資料私隱的角度來看，風險評估須考慮的因素包括：

- (i) 用來訓練人工智能模型的資料或輸入人工智能系統用作決策的資料的准許用途，當中須考慮《私隱條例》的規定，尤其是保障資料第3原則<sup>13</sup>；
- (ii) 用於訓練人工智能模型或人工智能系統運作的資料（尤其是個人資料）的數量<sup>14</sup>；
- (iii) 所涉資料的敏感程度。一般被視為較敏感的資料包括生物辨識資料、健康資料，以及弱勢群體（例如兒童）的個人資料；
- (iv) 所涉資料的質素，當中須考慮其來源、可靠性、真實性、準確性、一致性、完整性、相關性及可用性<sup>15</sup>；
- (v) 在開發或使用人工智能時的個人資料保安，當中應考慮到個人資料如何在機構的科技生態系統以及人工智能系統中轉移<sup>16</sup>；及
- (vi) 私隱風險（例如過度收集、濫用或外洩個人資料）出現的可能性及其潛在傷害的嚴重程度。

從更寬的道德標準角度來看，如果使用人工智能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，風險評估須考慮的因素亦應包括：

- (i) 人工智能系統對受影響個人及社會大眾的潛在影響（包括益處和傷害）；
- (ii) 人工智能系統的影響出現的可能性，以及其嚴重程度和持續時間；及

<sup>13</sup> 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。有關保障資料第3原則的詳情，請參閱附錄B。

<sup>14</sup> 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須屬足夠但不超乎適度。有關保障資料第1原則的詳情，請參閱附錄B。

<sup>15</sup> 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。有關保障資料第2原則的詳情，請參閱附錄B。

<sup>16</sup> 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。有關保障資料第4原則的詳情，請參閱附錄B。

(iii) 降低風險的緩減措施（包括技術性與非技術性措施）的足夠程度。

對個人來說，有關影響可能是影響其法律權利、人權（包括私隱權）、就業或教育前途，以及使用服務的資格等。如果人工智能系統很可能對持份者（尤其是個人）造成重大影響，有關系統會被視為高風險。

如果人工智能系統很可能對持份者（尤其是個人）造成重大影響，有關系統會被視為高風險。

圖 6 風險評估須考慮的因素（非詳盡無遺）



## 4.2.2 決定人為監督的程度

### 主要原則：人為監督

風險評估的主要目的是讓機構採取適當的風險管理措施，從而減低已識別的風險。機構在開發及使用人工智能時應採取風險為本的方式。因此，所採取的緩減風險措施（包括人為監督）的類別和程度，應與已識別的風險及風險程度相符合和相稱。機構應把無法消除的剩餘風險告知人工智能系統的使用者。無論如何，人工智能系統的剩餘風險應降至可接受水平。如果剩餘風險已在合理地切實可行的情況下減至最低，並與人工智能系統可為持份者帶來的益處相稱，則被視為可接受。

人為監督是減低使用人工智能的風險的主要措施。人工智能系統的風險評估結果會顯示使用人工智能系統時所需的適當人為監督的程度。無論如何，對人工智能所作的決策負上最終責任的應是人類行動者。

無論如何，對人工智能所作的決策  
負上最終責任的應是人類行動者。

一般來說，風險較高（即很可能對持份者造成重大影響）的人工智能系統，須有較高程度的人為監督：

- (i) 高風險的人工智能系統應以「人在環中」(human-in-the-loop)方式進行人為監督。在這模式中，人類行動者在決策過程中保留著控制權，以防止人工智能出錯或作出不當決定。
- (ii) 沒有真正風險或低風險的人工智能系統可採取「人在環外」(human-out-of-the-loop)方式進行人為監督。在這模式中，人工智能系統可以在沒有人為介入下作出決定，以達致完全自動化決策。

- (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，而「人在環中」的方式又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，人類行動者監督人工智能系統的運作，在有需要時才介入。

以下使用人工智能的例子屬於風險較高、須有較高程度的人為監督：

- (i) 使用生物辨識資料（例如臉容識別、聲紋識別和步態識別）實時識別個人，並可導致對有關個人採取不利行動；
- (ii) 招聘、工作表現評核或終止僱傭合約；
- (iii) 公共機構評估個人享用社會福利或公共服務的資格；及
- (iv) 評估個人的信用可靠程度，以便在提供貸款或其他金融服務方面作出自動化決策。

圖 7 風險為本的人為監督



### 4.3 人工智能模型的開發及人工智能系統的管理

透過機器學習而開發人工智能模型涉及幾個步驟，包括(1) 收集數據，(2) 準備數據，(3) 選擇機器學習模型（例如監督學習模型<sup>17</sup>及無監督學習模型<sup>18</sup>）及演算法，(4) 把訓練數據提供予機器學習演算法作分析以開發人工智能模型，及(5) 測試、評估及調校人工智能模型。訓練數據的數量和質素，以及所用的機器學習模型和演算法，均會對人工智能模型的準確性和可靠性有重大影響。

人工智能模型被採納使用後，或會繼續學習和演變。人工智能系統的操作環境亦會不斷轉變。因此機構在採用人工智能模型後，仍需繼續監察和檢視情況，並向使用者提供支援，以確保人工智能系統保持有效、相關和可靠。下文會建議開發人工智能模型及管理人工智能系統的措施。

圖 8 開發人工智能模型的過程



17 監督學習 (supervised learning model) 是一種機器學習，利用已標籤的數據集來訓練人工智能模型，令已受訓練的人工智能模型能夠把數據分類和作出預測。例如，把貓的相片標籤為「貓」，然後提供予機器學習演算法以訓練人工智能模型。已受訓練的人工智能模型便可識別相片中的貓。透過監督學習而開發的人工智能模型會提供較準確的預測。不過，它們未必能給予新的構想。

18 無監督學習 (unsupervised learning model) 涉及以機器學習演算法來分析無標籤的數據集，讓演算法自行從數據集發現規律和得出構想，無需人為介入。例如，網上零售商可利用無監督學習來分析其顧客的網上行為，以找出不同顧客的喜好和需要。透過無監督學習開發的人工智能模型可提供較有用的新構想，不過人工智能模型的透明度可能較低，其作出的預測的可解釋性亦會較低。

### 4.3.1 為人工智能準備數據

#### 主要原則：數據私隱 / 公平

人工智能在訓練及決策階段均使用數據，以找出規律、作出推斷、建議或決策。在這過程中時常會涉及個人資料。有效的數據管治不單保障個人的個人資料私隱，亦可確保數據質素良好，這對人工智能系統的公平性至為重要。管理不善的數據會引致「廢料進，廢品出」的情況，並對人工智能系統產生的結果有不利的影響。

數據質素對人工智能系統的公平性至為重要。

機構利用數據訓練人工智能模型之前，應採取下述步驟準備數據集：

(i) **機構必須採取措施，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<sup>19</sup>；
- 不要將個人資料用於與原本的收集目的不相符的目的，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<sup>20</sup>；
- 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<sup>21</sup>；
- 採取所有切實可行的步驟，確保個人資料安全<sup>22</sup>；及
- 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<sup>23</sup>。

19 見保障資料第1原則

20 見保障資料第3原則

21 見保障資料第2(1)原則

22 見保障資料第4原則

23 見《私隱條例》第26條及保障資料第2(2)原則

(ii) 在開發及使用人工智能時減少所使用的個人資料數量，可減低私隱風險。機構應採取下述措施和方法（如適用），把收集及使用的個人資料減至最少：

- 只收集與有關人工智能要達致的特定目的相關的數據，刪除含有個人特徵並與特定目的無關的數據；
- 使用匿名化<sup>24</sup>、假名化<sup>25</sup>或合成<sup>26</sup>數據來訓練人工智能模型；
- 在發放數據集供人工智能模型訓練之用前，對數據集應用「差分私隱<sup>27</sup>」技術；
- 採用聯合學習<sup>28</sup>方式訓練人工智能模型，以避免不必要地共享不同來源的訓練數據；及
- 如人工智能系統內的個人資料不再需要用於人工智能的開發及使用，應刪除有關資料。

(iii) 機構應管理用以訓練人工智能模型的數據質素，尤其是當人工智能系統所作的決定會對個人造成重大影響。有關數據應該是可靠、準確、完整、相關、沒有不義的偏見和非法的歧視。因此，機構應考慮下述事宜：

- 了解數據的來源、可靠性、真實性、準確性、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，並在有需要時移除或取代有關數值；
- 在使用數據訓練人工智能模型前，測試有關數據的公平性；及

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份的數據集。由於匿名化數據不能用來識別個人，它不是個人資料。

25 在假名化數據 (pseudonymised data) 內，可識別個人身份的資料已被移除，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份。假名化數據仍屬個人資料，因為在額外資料的輔助下，個人仍然可以被間接地識別出來。

26 合成數據 (synthetic data) 是指人工生成的數據集，與真實人士無關，因此應該沒有私隱風險。

27 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更強。

28 聯合學習 (federated learning) 是指由多個獨立的電腦系統合作開發人工智能模型。過程是首先由獨立的系統利用其系統內的數據各自開發人工智能模型。此舉可避免把訓練數據傳往中央數據庫，從而減低私隱及數據安全的風險。只有經訓練的人工智能模型才會從各自的系統輸出，再進一步共同開發一個整合的人工智能模型。

例如，如果某些組別人士的代表人數不足或超出比例，便可能令數據集存在不義的偏見。要處理這些問題，可採用不同的抽樣方法來重新平衡各類別的樣本的分佈。抽樣方法的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）。

- 指派人員定期檢視及更新訓練數據集，以確保數據質素良好。
- (iv) **機構應妥善記錄處理數據的情況**，以確保數據的質素和保安能保持良好，以及符合《私隱條例》的規定。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、篩選及在機構內轉移；
  - 數據的儲存地方；及
  - 如何維持數據質素。

### 4.3.2 人工智能模型的開發

主要原則：透明度與可解釋性 / 可靠、穩健及安全

在備妥數據後，機構可應用機器學習演算法，分析訓練數據，以開發人工智能模型。機構應了解不同機器學習演算法的特點，從而揀選符合其需要的演算法，當中要考慮的因素包括人工智能系統所輸出的分析結果要達致的準確程度及可解釋程度。

機構應了解不同機器學習演算法的特點，  
揀選符合其需要的演算法。

除了揀選適合的機器學習演算法，機構亦應考慮採取下述措施，以改善人工智能系統：

- (i) 對人工智能系統進行嚴格測試，確保系統可靠、穩健及公平，例如：
  - 將人工智能的決定與由人類或傳統非人工智能模型所作的決定互相比較；
  - 使用邊緣案例、未見過的數據<sup>29</sup>或有可能出現的惡意輸入，測試人工智能模型；及
  - 對人工智能系統進行可重複性及可再現性<sup>30</sup>的測試；
- (ii) 實施措施減低人工智能系統被輸入惡意資料或訓練數據的風險；
- (iii) 設立多重的緩衝層，以阻截人工智能系統的不同層面或模組所發生的錯誤或故障；
- (iv) 制定可以對人工智能系統的運作進行人為監督及介入的控制措施；
- (v) 制定保障人工智能系統及數據免受攻擊和外洩的保安措施；

<sup>29</sup> 未見過的數據指未曾用於訓練人工智能模型的數據集。它是用來測試人工智能模型的表現，因此亦稱為測試數據。

<sup>30</sup> 可再現性指在使用相同的數據集或預測方法時，人工智能系統是否產生相同的結果。可再現性對評估人工智能系統的可靠性十分重要。

- (vi) 制定應變計劃，在有需要時迅速暫停人工智能系統及啟動後備解決方案；
- (vii) 建立機制，確保人工智能系統的運作具足夠的透明度，讓使用者可以解釋系統輸出的結果；及
- (viii) 建立機制，提升人工智能系統的可追溯性<sup>31</sup>和可審核性，例如在人工智能系統運作時，自動記錄運作情況（即日誌）。

### 4.3.3 人工智能系統的管理與監察

主要原則：可靠、穩健及安全 / 人為監督

機構應不斷監察及檢視人工智能系統，因為應用人工智能系統的風險因素，包括訓練數據的相關性及人工智能模型的可靠性，會隨著時間而改變。這會影響人工智能系統的可靠性、穩健性及安全性。持續監察及檢視人工智能系統的方式會視乎風險程度而有所不同。高風險的人工智能系統需要較頻密和嚴格的監察及檢視。

因此，機構應考慮採取下述檢視機制：

- (i) 對人工智能系統的風險評估、設計、開發、測試及使用妥善地記錄存檔；
- (ii) 如人工智能系統在功能或運作上有重大改變，或規管或科技環境出現重大變化<sup>32</sup>，須重新評估人工智能系統的風險，以識別及應對新的風險；
- (iii) 定期檢視人工智能模型，確保模型的運作及表現符合預期；
- (iv) 定期以新數據調校及再訓練人工智能模型；
- (v) 確保在考慮人工智能系統的風險概況後，對人工智能系統訂立適當程度的人為監督；

31 可追溯性是指能夠記錄人工智能系統的開發及使用，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。確保可追溯性有助提升可審核性。

32 簡單的電腦保安修補及修正電腦程式錯誤通常不會觸發重新評估人工智能系統的風險的需要。

人為監督的目的應是避免及降低人工智能對個人造成的風險。進行人為監督的人員應能：

- 充分了解人工智能系統的能力和限制；
- 對可能過份依賴人工智能輸出的結果的狀況（即自動化偏見）保持警覺；
- 正確地解釋人工智能輸出的結果；
- 在人工智能輸出的結果出現異常時，不理會、撤銷或推翻結果；及
- 在適當時介入及中斷人工智能系統的運作。

(vi) 在人工智能系統由開發、使用、監察以至終止的整個生命周期維持穩健的保安措施；

(vii) 為人工智能系統的使用者持續提供運作上的支援及反饋途徑；及

(viii) 定期評估宏觀的科技環境，以識別與機構現時的科技生態系統的差距，並在有需要時調整人工智能策略及管治架構。

機構應定期進行內部審核，以確保人工智能的開發及使用遵從機構相關政策的規定，以及與人工智能策略保持一致。內部審核的結果應向機構的高級管理層及管治組織（例如審核委員會）匯報。

圖 9 人工智能系統的開發及管理



#### 4.4 與持份者的溝通及交流

##### 主要原則：透明度與可解釋性

機構應讓持份者知道它們在使用人工智能，以展示機構奉行上述三項數據管理價值，同時避免或減少使用人工智能可能造成的傷害。因此，開發及使用人工智能系統的機構應與持份者（尤其是個別的消費者及規管者）有效地溝通及交流。有效的溝通亦是建立信任所必需的。

因此，在與持份者溝通方面，機構應考慮採取下述步驟：

- (i) 除非有關情況和環境已清楚顯示機構正在使用人工智能系統，否則機構應明確地向個人披露它們有使用人工智能系統；
- (ii) 機構應就有關在產品或服務中使用人工智能系統的目的、益處、限制及效果提供足夠的資訊，除非有關披露會對商業敏感資訊造成損害；及
- (iii) 機構應披露人工智能系統的風險評估結果，除非有關披露會對商業敏感資訊造成損害。

至於可能對個人造成重大影響的人工智能系統，機構亦應向個人提供途徑讓他們改正不準確之處、作出反饋、尋求解釋、要求人為介入及 / 或退出使用人工智能（如適用）。

如情況可行，在為人工智能作出或輔助作出的決策提供解釋時，可包括下述資訊<sup>33</sup>：

- (i) 人工智能如何參與決策過程及其參與的程度，包括所採用的人工智能系統主要負責的工作，以及人類行動者的參與情況；
- (ii) 自動化或人工智能輔助決策過程所使用的數據類別，以及為何這些類別的數據被視為相關和必需；
- (iii) 自動化決策過程所使用的個人概況是如何彙編的，包括分析中有否使用任何統計數據，以及為何有關個人概況與自動化決策過程有關聯；及
- (iv) 作出有關自動化決策和最終決策（如兩者不同）的主要因素。

33 有關更多如何有意義地解釋人工智能所作的自動化決策的建議，機構可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《解釋人工智能決策》指引。

圖 10 與持份者的溝通及交流



機構在決定披露的資訊類別及詳細程度時，其中應考慮有關持份者理解資訊的能力、他們的需要，以及有關披露會否對人工智能系統的保安和合法目的造成不利影響。例如，披露予普通消費者的資訊不應過於技術性，否則他們可能不明白。對於用來偵測顧客欺詐行為的人工智能系統，有關機構毋須披露人工智能系統用以判辦欺詐行為的指標，以免顧客有機會「欺騙」該系統。

與持份者（尤其消費者）的通訊應以淺白語言、清楚易明的方式書寫，並設法讓持份者知悉。有關資訊亦可以納入機構的私隱政策內。

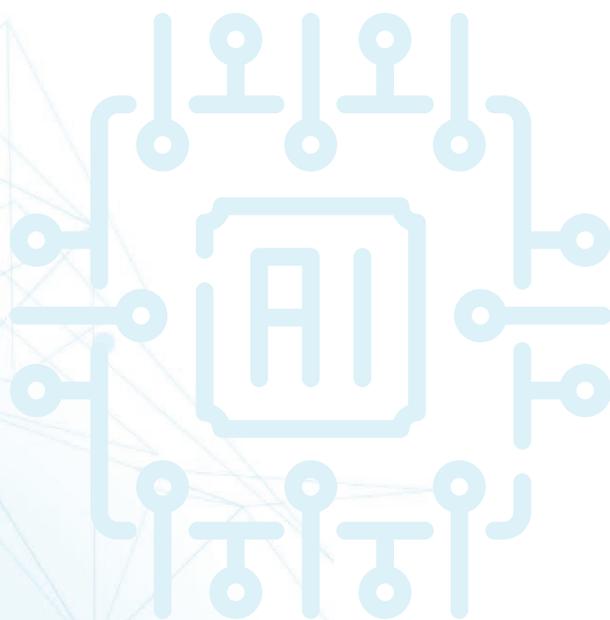
與持份者的通訊應以淺白語言、清楚易明的方式書寫，  
並設法讓持份者知悉。

# 5 由第三方提供的人工智能系統



本指引主要為自家開發人工智能系統的機構提供建議。至於聘請第三方承辦商開發人工智能系統或購買現成的人工智能系統的機構，它們應採取適當的步驟，確保依從本指引所建議的原則及措施。例如，機構可要求第三方承辦商在開發人工智能時依從本指引的建議。機構在使用現成的人工智能系統之前，亦可測試系統的可靠性、穩健性和公平程度。

即使人工智能系統是由第三方開發，使用有關系統的機構仍須為系統所作的決策以及遵從《私隱條例》的規定和更廣寬的道德原則負上責任。



# 6 結語



「人工智能有潛質為社會帶來巨大的益處，但先決條件是要負責任地使用它<sup>34</sup>。」當世界各地的監管機構正考慮應否以及如何利用法律和規例明確直接地規管人工智能的開發及使用之時，開發及使用人工智能的機構必須遵從保障個人資料的相關法律，以及應實踐良好的數據道德標準。因此，我們促請機構遵從本指引所建議的價值、原則和措施。

總括來說，在數據主導的經濟環境中，信任是十分重要的。本指引所建議的價值、原則和措施正是為機構提供獲取顧客、持份者及社會大眾信任的方法。



<sup>34</sup> 世界經濟論壇創辦人兼執行主席克勞斯·施瓦布(Klaus Schwab)教授。摘自世界經濟論壇的新聞稿《世界經濟論壇成立新的環球聯盟推動使用負責任的人工智能》(2021年1月28日)

# 附錄 A - 自我評估核對清單

## 人工智能策略及管治

	問題	答案 (有/沒有)	所需的進一步行動
1	貴機構在開發及使用人工智能前有沒有制定人工智能策略？		
2	貴機構有沒有就合乎道德標準地設計、開發及使用人工智能制定具體的內部政策和程序？		
3	貴機構有沒有成立人工智能管治委員會（或類似組織）以監督人工智能系統由開發、使用，以至終止的整個生命周期？		
4	人工智能管治委員會（或類似組織）有沒有： <ul style="list-style-type: none"> <li>來自不同專業領域及部門的成員就人工智能的開發及使用互相合作？</li> <li>高級管理人員監督其運作？</li> </ul>		
5	貴機構有沒有為開發及使用人工智能的人員訂明清晰的角色及責任？		
6	貴機構有沒有在財政和人力上提供足夠的資源以開發及使用人工智能？		
7	貴機構有沒有為開發及使用人工智能的人員提供與其職位相關的培訓？		
8	貴機構有沒有定期為所有相關人員安排活動，加強他們對使用人工智能的認識？		

## 風險評估及人為監督

	問題	答案 (有 / 沒有)	所需的進一步行動
1	貴機構有沒有在開發及使用人工智能之前進行風險評估？		
2	貴機構的風險評估有沒有考慮個人資料私隱的風險及人工智能系統的道德方面的影響？		
3	風險評估結果有沒有由人工智能管治委員會（或類似組織）檢視及認可？		
4	貴機構有沒有在考慮過人工智能系統的風險概況後，對人工智能系統訂立適當程度的人為監督及其他緩減風險的措施？		

## 人工智能模型的開發及人工智能系統的管理

	問題	答案 (有 / 沒有)	所需的進一步行動
<b>準備數據</b>			
1	貴機構有沒有採取步驟把使用的個人資料減至最少及確保遵從《私隱條例》的規定（例如使用匿名化或合成數據；了解個人資料的來源及准許的用途；檢查個人資料的準確性等）？		
2	貴機構在使用數據前，有沒有採取步驟確保數據的可靠性、真實性、準確性、一致性、完整性、相關性、公平性及可用性？		

	問題	答案 (有/沒有)	所需的進一步行動
<b>人工智能模型的開發</b>			
3	貴機構在使用機器學習演算法之前，有沒有評估其特點？		
4	貴機構有沒有對人工智能模型進行嚴格測試，確保模型可靠、穩健及公平？		
5	貴機構有沒有制定足夠的風險緩減措施，包括人為監督，以應付使用人工智能系統時可能出現的錯誤或故障？		
6	貴機構有沒有制定足夠的保安措施，保障人工智能系統免受攻擊？		
7	貴機構有沒有制定應變計劃，在有需要時暫停人工智能系統及啟動後備解決方案？		
<b>管理與監察</b>			
8	貴機構有沒有對人工智能系統的數據處理、風險評估、設計、開發、測試及使用妥善地記錄存檔？		
9	貴機構有沒有制定計劃當人工智能系統在功能或運作上有重大改變，或規管或科技環境出現重大變化時，重新評估人工智能的風險？		
10	貴機構有沒有定期檢視、調校及再訓練人工智能模型？		
11	貴機構有沒有根據評估的風險程度對人工智能系統訂立適當程度的人為監督？		
12	貴機構有沒有為人工智能系統的使用者提供運作上的支援及反饋途徑？		

	問題	答案 (有/沒有)	所需的進一步行動
13	貴機構有沒有在人工智能系統由開發、使用、監察以至終止的整個生命周期實施適當的保安措施？		
14	貴機構有沒有計劃定期評估宏觀的科技環境，以識別與機構現時的科技生態系統的差距？		
15	貴機構有沒有定期進行內部審核，確保人工智能的開發及使用遵從內部政策的規定？		

## 與持份者的溝通及交流

	問題	答案 (有/沒有)	所需的進一步行動
1	貴機構有沒有清楚明確地向個別的消費者披露機構使用人工智能系統？		
2	貴機構有沒有告知個別的消費者在產品或服務中使用人工智能系統的目的、益處及效果？		
3	貴機構有沒有披露人工智能系統的風險評估結果（如適合）？		
4	貴機構有沒有向個人提供退出使用人工智能的途徑（如可能）？		
5	貴機構有沒有向個人提供途徑讓他們改正不準確之處、作出反饋、尋求解釋以及要求人為介入（如可能）？		
6	與持份者的通訊有沒有以淺白語言、清楚易明的方式書寫？		

## 附錄 B - 《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，資料使用者只可以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，但可不超乎適度。

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，而保留時間不超過達致原來目的實際所需。《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 C - 參考資料

- 環球私隱議會，《人工智能的倫理道德與資料保障宣言》(2018)
- 歐盟委員會 – 人工智能獨立高級專家組，《可信賴的人工智能的道德準則》(2019)
- 香港金融管理局，《應用人工智能的高層次原則》(2019)
- 日本內閣府，《以人為本的人工智能社會原則》(2019)
- 經濟合作與發展組織，《經合組織理事會有關人工智能的建議》(2019)
- 環球私隱議會，《開發及應用人工智能的體現問責決議》(2020)
- 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》(第2版)(2020)
- 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、世界經濟論壇，《人工智能管理模範框架指南 – 機構的實施與自我評估指引》(2020)
- 英國資訊專員辦公室，《人工智能及資料保障指引》(2020)
- 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)
- 聯合國教科文組織，《人工智能道德建議書草案文本初稿》(2020)
- 歐盟委員會，《關於制定人工智能統一規則的條例提案》(2021)
- 香港特區政府資訊科技總監辦公室，《人工智能道德框架》(2021)





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



查詢熱線 Enquiry Hotline : (852) 2827 2827

傳真 Fax : (852) 2877 7026

地址 Address : 香港灣仔皇后大道東248號大新金融中心13樓1303室  
Room 1303, 13/F, Dah Sing Financial Centre,  
248 Queen's Road East, Wanchai, Hong Kong

電郵 Email : [communications@pcpd.org.hk](mailto:communications@pcpd.org.hk)



私隱公署網頁  
PCPD website  
[pcpd.org.hk](http://pcpd.org.hk)



下載本刊物  
Download  
this publication



本刊物使用署名4.0國際(CC BY 4.0)的授權條款，只要你註明原創者為香港個人資料私隱專員，便可自由分享或修改本刊物。詳情請瀏覽[creativecommons.org/licenses/by/4.0/deed.zh](http://creativecommons.org/licenses/by/4.0/deed.zh)。

This publication is licensed under Attribution 4.0 International (CC By 4.0) licence. In essence, you are free to share and adapt this publication, as long as you attribute the work to the Privacy Commissioner for Personal Data, Hong Kong. For details, please visit [creativecommons.org/licenses/by/4.0](http://creativecommons.org/licenses/by/4.0).

#### 免責聲明 Disclaimer

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。

The information and suggestions provided in this publication are for general reference only. They do not serve as an exhaustive guide to the application of the law and do not constitute legal or other professional advice. The Privacy Commissioner makes no express or implied warranties of accuracy or fitness for a particular purpose or use with respect to the information and suggestions set out in this publication. The information and suggestions provided will not affect the functions and powers conferred upon the Privacy Commissioner under the Personal Data (Privacy) Ordinance.