僱員使用生成式AI的指引清單

PCPD
H K

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

引言

因應生成式人工智能（生成式AI）的快速發展，個人資料私隱專員公署（私隱專員公署）制定此清單，旨在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《個人資料（私隱）條例》有關處理個人資料的相關規定，從而促進人工智能（AI）能在香港的安全及健康發展。作為良好的行事方式，機構應視乎自身情況制定與其公司價值觀及使命一致的政策或指引，並定期檢視及更新相關政策或指引，以反映運作上及技術上的改變。

此清單所指的「資訊」一詞，主要指個人資料；然而，視乎機構所處理的資訊性質及需要，亦可泛指一般資訊或資料。

生成式AI工具的例子

香港機構經常使用的生成式AI工具包括：

- 聊天機械人；
- 光學字符識別（Optical Character Recognition）；
- 文字／圖像／影片／語音生成器；
- 文件／簡報生成器；及
- 語音轉文字工具等。

視乎機構如何應用生成式AI工具，僱員取用這些工具的方法會有所不同。例如，這些工具可被安裝在僱員的電腦上運行（不論是否連接互聯網）、託管在機構的伺服器（例如內聯網）、透過應用程式介面（Application Programme Interface）取用或透過互聯網（例如網頁版工具）取用。

僱員使用生成式AI的政策或指引的建議內容

範圍

☑ **獲准使用的工具：**清晰訂明機構內准許使用的生成式AI工具及應用程式。這些工具可能包括：

- 公眾可用的生成式AI工具或應用程式；及／或
- 內部開發的生成式AI工具或應用程式。

保障資料小貼士💡：如使用公眾可用的生成式AI工具或應用程式，商業授權的版本可能比一般版本提供更多個人資料私隱及保安保障。如使用內部開發的生成式AI工具，將這些工具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，機構一般可掌握更多資料保安的控制權（例如要求所輸入及輸出的資料均儲存在僱員的裝置或機構的伺服器上）。然而，機構應評估它們是否有足夠專業知識及資源以安全地營運及保護內部伺服器上的系統。

☑ **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；及／或
- 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。

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，以及禁止輸入的資訊種類（例如：個人資料[1]、機密資料、專有資料或受版權保護的資料）[2]。

保障資料小貼士💡：如機構准許僱員輸入個人資料至生成式AI工具，私隱專員公署建議機構在可行及適當的情況下，指示僱員將個人資料匿名化，並就如何在輸入前將個人資料匿名化或「淨化」提供清晰指示。

☑ **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[3]，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，並根據機構的資料保留政策刪除生成式AI工具所生成的資訊，包括僱員所用的資訊。

☑ **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，包括有關處理個人資料及資訊保安的政策。

合法及合乎道德的使用及預防偏見

☑ 訂明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，以確保AI所生成的結果符合機構的道德價值觀及標準。

- **準確度及核實：**強調僱員需要核實AI所提供的資訊，包括進行校對及查核事實，以確保資訊是準確和最新的。
- **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，並訂明需要遵循的更正及報告機制。
- **加上水印／標籤：**提供清晰指引，說明應何時及如何在AI生成結果上加上水印或標籤。

1《個人資料（私隱）條例》保障資料第3原則訂明，如無有關資料當事人的訂明同意，個人資料不得用於在收集該資料時擬將該資料使用的目的以外的任何目的。

2 機構在衡量獲准輸入的資訊種類及數量時，應考慮獲准使用的生成式AI工具的保安及私隱保障水平。

3 見註1。

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，例如：

- 辦公室電腦；
- 工作手提電話；及／或
- 平板電腦。

保障資料小貼士💡：私隱專員公署建議機構要求僱員在工作裝置上使用生成式AI工具時，只限於與工作相關的用途。

☑ **獲准許使用者：**訂明可以使用生成式AI工具的僱員，例如有工作上的需要並曾接受相關培訓的僱員，以及是否需要事先獲得批准才可使用。

☑ **穩健的用戶憑證：**要求僱員使用獨特且高強度的密碼及多重認證。

☑ **保安設定：**要求僱員在生成式AI工具輸入工作相關的資料時，保持嚴格的保安設定。

保障資料小貼士💡：不在生成式AI工具上儲存或不與其供應商分享提示詞(prompts)，有助將資料保安事故及僱員的行為剖析風險降至最低。

☑ **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，包括以下事故：

- 涉及AI的資料外洩事故；
- 未獲授權下輸入個人資料；
- 異常的輸出結果；及／或
- 可能涉及違法的輸出結果。

保障資料小貼士💡：機構可參考私隱專員公署發布的《資訊及通訊科技的保安措施指引》[4]及《資料外洩事故的處理及通報指引》[5]，了解更多有關加強資料保安及處理資料外洩事故的貼士。

違反政策或指引

☑ 訂明僱員違反使用生成式AI政策或指引可引致的後果。

保障資料小貼士💡：機構可參考私隱專員公署發布的《人工智能(AI)：個人資料保障模範框架》[6]的建議，以制定生成式AI的管治架構及措施（例如，管理及持續監察僱員使用生成式AI工具的措施）。

4 請參閱：https://www.pcpd.org.hk/chinese/resources_centre/publications/files/guidance_datasecurity_c.pdf
5 請參閱：https://www.pcpd.org.hk/chinese/resources_centre/publications/files/guidance_note_dbn_c.pdf
6 請參閱：https://www.pcpd.org.hk/tc_chi/resources_centre/publications/files/ai_protection_framework.pdf

支援僱員使用生成式AI工具的實用貼士

☑ **透明度：**定期向僱員傳達政策或指引，以確保他們清楚了解是否獲准許使用及如何使用生成式AI工具，並及時告知僱員任何政策或指引的更新。

☑ **培訓及資源：**教育僱員如何有效及負責任地使用生成式AI工具，包括：

- 說明AI工具的能力及限制；
- 提供適當及安全地在工作上使用AI工具的實務建議及例子；及／或
- 鼓勵僱員閱讀AI工具的私隱政策、使用條款及其他處理資料的政策，以了解個人資料會如何被收集、儲存、使用及分享。

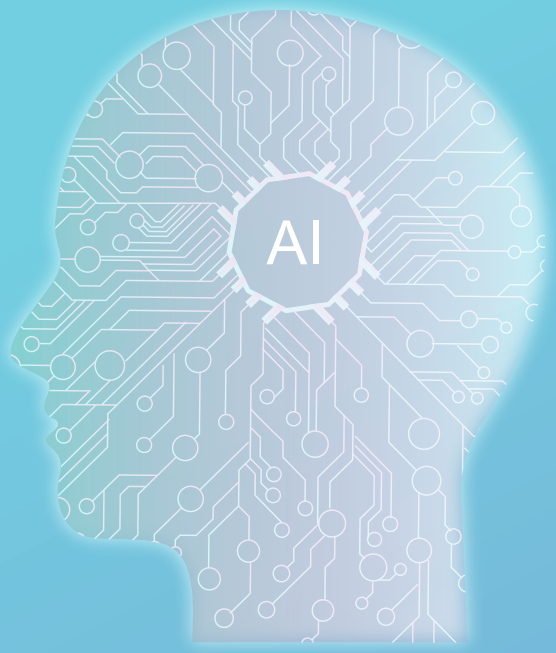
> **保障資料小貼士**💡：機構可鼓勵僱員閱讀私隱專員公署發布的《使用AI聊天機械人「自保」十招》[7]。

☑ **委派支援隊伍：**委派指定的支援隊伍協助在工作上使用生成式AI工具的僱員。除技術支援外，支援隊伍亦應能夠解答僱員有關政策或指引的任何疑慮。

> **保障資料小貼士**💡：支援隊伍可包括負責協助機構遵從資料保障法律的人員（例如資料保障主任）及負責機構的AI管治的人員（例如AI管治委員會或類似團體的成員）。

☑ **反饋機制：**建立渠道讓僱員提供在工作上使用生成式AI工具的反饋，以協助機構識別可以改進的地方，以及根據情況更新政策或指引。

7 請參閱：https://www.pcpd.org.hk/chinese/resources_centre/publications/files/ai_chatbot_leaflet.pdf



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

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。

二零二五年三月