使用生成式AI, 會有哪些個資風險？如何防免？

文:黃耀平（認證法律人） · 基本人權·政府 · 2025-03-14

本文

從ChatGPT掀起風潮開始，許多公司利用生成式AI提供類似的對話機器人服務（其他例如Gemini、Claude等），由於這類服務需要電腦的高速運算，一般用戶的手機或電腦速度又不夠快，因此公司會把用戶說的話，或者提供的資料透過網路，送到公司的機房，經過高速運算以後，再把結果發送給用戶。對於用戶和公司雙方而言，個資傳送的過程愈複雜、經過愈多手，風險就愈高，例如：網路安全機制不佳，被人輕易入侵，或者內部控管不當，員工不慎外洩等問題。

一、與生成式AI對話，用戶的個資會面臨什麼風險？

（一）什麼是個資？

個資，是個人資料的簡稱（本文簡稱「個資」），法律上是指自然人的姓名、出生年月日、身分證ID、護照號碼、特徵、指紋、婚姻、家庭、教育、職業、病歷、醫療、基因、性生活、健康檢查、犯罪前科、聯絡方式、財務情況、社會活動及其他得以直接或間接方式識別該個人的資料[1]。

所以，只要可以「直接或間接識別個人的資料」就是個資，不以上述舉例的種類為限，例如：信用卡卡號，搭配持卡人的簽名[2]，或是健身房的會籍，搭配會員姓名、聯絡電話等[3]，是法條列舉種類以外而可以連結到個人的資料，都會是個資。

（二）個資洩漏的種類、方式

前面提到的自然人姓名、出生年月日、婚姻、家庭、職業等個資，都是在與生成式AI對話時，可能提供的資料，例如：上傳自己的醫療資訊要求提醒按時服藥，或上傳客戶的財務數據請求提供理財建議等，在與生成式AI對話享受便利之際，應注意提供這些個資，可能有遭不法利用、洩漏等風險。

使用生成式AI，還有一種特殊的外洩個資的方式，類似「說溜嘴」的現象，例如：對話機器人可能會將A用戶的對話，或者提供的資料（例如：簡報、開會內容、醫療資料），輸入它的系統加以訓練，用來改善系統，如果管理不慎，就可能將A的資料提供給B用戶（就好像不小心說溜嘴一樣）。根據國外的新聞報導，曾有對話機器人將A用戶對話內容，提供給B用戶的事件[4]。

（三）如何防止洩漏個資？

為了避免與對話機器人交談的過程中洩漏個資，如果是親友等其他人的個資，在沒有經過本人同意前，不要向生成式AI提供[5]；自己的個資，也建議如非必要不提供，避免個資被生成式AI以「說溜嘴」等方式提給其他用戶。

二、開發生成式AI的公司，如何保存用戶的個資？

（一）個資的保存義務

如果您是提供生成式AI的企業經營者，手握用戶個資，在服務時除了應明確告知用戶，關於其個資的使用目的以外[6]，並應注意避免生成的內容不慎外洩個人資料。個人資料保護法27條要求個資保有者應採行「適當的安全措施」，防止個人資料被竊取、竄改、毀損、滅失或洩漏[7]。個資如不慎外洩，造成用戶損害，有可能要負損害賠償責任[8]。

（二）公司可以採取哪些適當的安全措施？

所謂保存個資的「適當的安全措施」，會依照公司所保有的個資數量、公司的規模以及資源等等，來判斷採取的措施，究竟應達到何種程度[9]。常見的網路安全措施，例如：設定帳號、高強度密碼、多重認證機制、設定存取層級、防毒軟體、防火牆、系統定時更新、系統漏洞補強、門禁管理、定期檢討等等，還有最重要的「去識別化」，都是防止外洩的安全措施。

如果公司沒有採取適當的安全措施，可能會面臨新臺幣（下同）2萬～200萬元的罰鍰，並被主管機關要求限期改正，期限內沒改的話，還會再被按次罰15萬～1,500萬元[10]。

（三）去識別化仍應避免被還原

企業經營者將保有的個資「去識別化」是防止外洩的好方法，因為當個人的資料已無法識別，就不再是受法律保障的個資了，例如：將姓名改成黃○○、電話號碼改成09xx－xxxxxx；適當地去識別化[11]，能保障客戶的隱私，也能避免因為個資外洩而負上賠償責任。

然而，在隱去個資的過程，應注意資料是否能被還原，業界一個著名的例子，很適切地說明了這個風險：知名的網飛（Netflix）公司，為了準確預測觀眾的喜好，他們舉辦了預測競賽，預測最準確者能獲得高額獎金，Netflix提供其用戶的去識別化資料給參賽者使用，包含觀看的影片、觀看時間、對影片的評分等資料。雖然上述資料已做過去識別化的處理，然而德州大學奧斯汀分校的研究團隊對照Netflix提供的資料以及其他網路電影資料庫後，成功將部分用戶的紀錄去匿名化，識別出其身分[12]，之後亦衍生法律糾紛。由此可知，去識別化的處理應該更加謹慎，如果能被輕易地逆向還原，就不是適當的去識別化。

綜上，如果您是生成式AI的使用者，應注意服務供應商是否值得信賴，如非必要不提供個人資料；如果您是企業經營者，應該在去識別化多花功夫，也要採取適當安全措施，以防止個資外洩。

註腳

[1] 個人資料保護法第2條第1款：「本法用詞，定義如下：一、個人資料：指自然人之姓名、出生年月日、國民身分證統一編號、護照號碼、特徵、指紋、婚姻、家庭、教育、職業、病歷、醫療、基因、性生活、健康檢查、犯罪前科、聯絡方式、財務情況、社會活動及其他得以直接或間接方式識別該個人之資料。」

[2] 國家發展委員會發法字第1100086051號（2021/11/25）：「查貴會函詢旨揭新聞畫面，涉及信用卡號及持卡人簽名等資料，因該等資料得藉由與其他資料對照組合而間接識別特定之個人，爰依上開個資法第2條第1款及同法施行細則第3條規定，該等資料係屬個人資料而有個資法之適用。」

[3] 臺灣新竹地方法院113年度訴字第156號刑事判決：「甲○○、乙○○均任職在WorldGym世界健身俱樂部竹北店……擔任教練，明知對於會員所留底之個人資料均應保密並依個人資料保護法相關規定辦理，且WorldGym世界健身俱樂部是會員制的俱樂部，課程須在有效的會籍狀況下進行，無法單買課程，二人為使非會員之丙○○能購買課程，竟意圖為自己或第三人不法之利益，共同基於行使偽造私文書，及違法利用個人資料之犯意聯絡，未經俱樂部會員丁○○之同意，而違法利用丁○○之姓名、身分證字號、聯絡方式等個人資料及會籍，由甲○○以電子手寫板，在『個人教練課程合約書』，填載丁○○之會員資料，並於『個人教練課程合約書』、『個人訓練課程約定事項』偽簽丁○○簽名共3枚，傳送予WorldGym世界健身俱樂部竹北店及丁○○以行使，二人並一同於櫃檯操作課程訂購等事宜，為不知情之非會員丙○○購買課程，足生損害於丁○○之隱私及資訊自主權。」

[4] Ian Krietzberg (2024), ChatGPT is leaking users' passwords, report finds, TheStreet.

[5] 參行政院及所屬機關（構）使用生成式AI參考指引第4點：「業務承辦人不得向生成式AI提供涉及公務應保密、個人及未經機關（構）同意公開之資訊，亦不得向生成式AI詢問可能涉及機密業務或個人資料之問題。但封閉式地端部署之生成式AI模型， 於確認系統環境安全性後，得依文書或資訊機密等級分級使用。」

[6] 個人資料保護法第8條第1項第2款：「公務機關或非公務機關依第十五條或第十九條規定向當事人蒐集個人資料時，應明確告知當事人下列事項：……二、蒐集之目的。」
個人資料保護法第19條第1項：「非公務機關對個人資料之蒐集或處理，除第六條第一項所規定資料外，應有特定目的，並符合下列情形之一者：……。」
個人資料保護法第20條第1項本文：「非公務機關對個人資料之利用，除第六條第一項所規定資料外，應於蒐集之特定目的必要範圍內為之。」

[7] 個人資料保護法第27條第1項：「非公務機關保有個人資料檔案者，應採行適當之安全措施，防止個人資料被竊取、竄改、毀損、滅失或洩漏。

[8] 個人資料保護法第29條：「
I 非公務機關違反本法規定，致個人資料遭不法蒐集、處理、利用或其他侵害當事人權利者，負損害賠償責任。但能證明其無故意或過失者，不在此限。
II 依前項規定請求賠償者，適用前條第二項至第六項規定。」
個人資料保護法第28條第2至6項：「
II 被害人雖非財產上之損害，亦得請求賠償相當之金額；其名譽被侵害者，並得請求為回復名譽之適當處

分。

III 依前二項情形，如被害人不易或不能證明其實際損害額時，得請求法院依侵害情節，以每人每一事件新臺幣五百元以上二萬元以下計算。

IV 對於同一原因事實造成多數當事人權利受侵害之事件，經當事人請求損害賠償者，其合計最高總額以新臺幣二億元為限。但因該原因事實所涉利益超過新臺幣二億元者，以該所涉利益為限。

V 同一原因事實造成之損害總額逾前項金額時，被害人所受賠償金額，不受第三項所定每人每一事件最低賠償金額新臺幣五百元之限制。

VI 第二項請求權，不得讓與或繼承。但以金額賠償之請求權已依契約承諾或已起訴者，不在此限。」

[9] 個人資料保護法施行細則第12條：「

I 本法第六條第一項但書第二款及第五款所稱適當安全維護措施、第十八條所稱安全維護事項、第十九條第一項第二款及第二十七條第一項所稱適當之安全措施，指公務機關或非公務機關為防止個人資料被竊取、竄改、毀損、滅失或洩漏，採取技術上及組織上之措施。

II 前項措施，得包括下列事項，並以與所欲達成之個人資料保護目的間，具有適當比例為原則：

一、配置管理之人員及相當資源。

二、界定個人資料之範圍。

三、個人資料之風險評估及管理機制。

四、事故之預防、通報及應變機制。

五、個人資料蒐集、處理及利用之內部管理程序。

六、資料安全管理及人員管理。

七、認知宣導及教育訓練。

八、設備安全管理。

九、資料安全稽核機制。

十、使用紀錄、軌跡資料及證據保存。

十一、個人資料安全維護之整體持續改善。」

[10]個人資料保護法第48條第2項：「非公務機關違反第二十七條第一項或未依第二項訂定個人資料檔案安全維護計畫或業務終止後個人資料處理方法者，由中央目的事業主管機關或直轄市、縣（市）政府處新臺幣二萬元以上二百萬元以下罰鍰，並令其限期改正，屆期未改正者，按次處新臺幣十五萬元以上一千五百萬元以下罰鍰。」

[11]在技術上除了「單純隱去」以外，還有較為複雜的資料處理的方法，例如編碼法、差分法等等，都值得手握眾多個資者進一步研究。例如關於差分法的好處，參照Adam McCormick, Amol Khanna (2024), Accelerating differential privacy deployment in the federal government, IAPP.

國內文獻，可參考翁清坤（2023），〈個人資料之去識別化與再識別化風險：法律之觀點〉，《臺大法學論叢》，第52卷第3期，頁619-739。

[12]Arvind Narayanan, Vitaly Shmatikov (2008), Robust De-anonymization of Large Sparse Datasets, 2008 IEEE Symposium on Security and Privacy (sp 2008), Oakland, CA, USA, pp. 111-125.

標籤

生成式AI， ChatGPT， 個人資料， 對話機器人， 去識別化