

法規名稱：臺北市政府警察局人工智慧使用規範

制(訂)定日期：民國 115 年 10 月 27 日

當次沿革：中華民國 115 年 10 月 27 日臺北市政府警察局北市警資字第 1153102521 號函訂定全文 7 點；並自 115 年 10 月 27 日生效

一、臺北市政府警察局（以下簡稱本局）為規範所屬各單位研發、採購、部署、維運或使用人工智慧（以下簡稱 AI）系統之行為，確保警政科技之應用符合人類自主、保障基本權利、維護資訊安全、資料治理與法遵合規，參照「人工智慧基本法」、「人工智慧風險分類框架」及「臺北市政府使用人工智慧作業指引」等規定，訂定本規範。

二、本局各單位使用 AI 進行各項警政研發應用、創新實驗環境、勤務部署及數據處理等行為時，應遵循本規範並確認附件一「臺北市政府警察局人工智慧應用檢核表」（以下簡稱檢核表）之項目是否均符合規定。

本局各單位辦理採購事項，應要求得標之事業、法人、團體或個人遵循本規範，並於招標文件及契約中強制約定承接廠商共同遵循。

三、本規範所稱人工智慧，依人工智慧基本法第三條，指具自主運行能力之系統，該系統透過輸入或感測，經由機器學習及演算法，可為明確或隱含之目標實現預測、內容、建議或決策等影響實體或虛擬環境之產出。

本局 AI 應用依功能及運用方式，包含下列類型；同一應用得兼具數種類型，其風險應依實際使用情境判斷：

- （一）分析式 AI：使用統計模型或自然語言處理技術來洞悉數據或資料之模式或規律。例如分析治安事件、刑案資料、一一〇報案紀錄、交通事故、交通違規或勤務紀錄，以掌握時空分布及趨勢。
- （二）預測式 AI：基於歷史數據來預測未來可能發生的事件、趨勢或結果。例如預測治安事件或交通事故發生風險、道路壅塞狀況、特定時段勤務需求，或提供警力及交通疏導資源配置之參考。
- （三）生成式 AI：依據使用者所輸入之提詞，生成文字、語音、圖案或影片等內容。例如依據警政法規、函釋、勤務規範及標準作業程序，草擬行政文書、會議紀錄、新聞稿、教育訓練教材、勤務提示或案件摘要，並提供內部警政知識問答服務。
- （四）代理式 AI：依據使用者所設定之目標，自主進行任務規劃、工具

調用、環境感知與反思優化，以獨立執行多步驟之複雜任務或流程。

有關人工智慧的架構與技術等詳細說明，請參閱附件二「人工智慧的應用開發」。

四、本局 AI 應用之風險分級管理框架，分為絕對禁止事項、高度風險、中度風險及低度風險等四個等級。各單位於評估導入 AI 應用方案前，應逐一核對檢核表，並經本局 AI 推動小組審查判定風險等級後，始得踐行相應之建置與審查程序。

（一）絕對禁止事項：

使用人工智慧涉及機敏性資料、公務上應保密資料或未經本局同意公開之資料，原則列為絕對禁止事項。但使用本局自建或不對外聯網之地端機房，或相同資安防護等級之自控機房，於確認系統安全後再依機密等級分級使用者，不在此限。以下為絕對禁止事項之使用情境：

1. 未經授權將機密資料、公務上應保密資料、犯罪偵查資料、警政機敏資料、民眾個人資料或其他未經本局同意公開之資料，輸入外部或未經核准之 AI 系統。
2. 未具法令依據或未符合個人資料保護法規定，使用 AI 蒐集、處理或利用民眾之個人資料、特種個人資料、生物特徵、行蹤軌跡或其他隱私資料。
3. 未經權責人員實質檢視、判斷及確認，直接以 AI 產出作成行政處分、交通舉發、案件派遣、勤務部署或其他影響民眾權利義務之決定。
4. 僅依 AI 之評分、分類、預測或辨識結果，認定特定人具有犯罪嫌疑、危險性或違法事實，並據以採取臨檢盤查、搜索、扣押、拘提、逮捕或其他公權力措施。
5. 使用 AI 偽造、變造、隱匿或不當生成證據、影像、錄音、筆錄、公文、系統紀錄或其他公務資料。
6. 未經核准，使 AI 自主操作警政系統、修改正式資料、對外傳送資訊或執行足以影響民眾權利義務之行為。
7. 其他違反法令、資訊安全、個人資料保護或本局內部管理規定之情形。

（二）高度風險之使用情境：

涉及民眾權利義務、隱私或個人資料，可能產生偏見歧視，或依前述評估可能造成嚴重危害者，應列為高度風險。以下為高度風險之使用情境：

1. 分析涉及個人資料或機敏性資料之刑案、犯罪情資、一一〇報案、車籍、戶籍、行蹤軌跡或其他警政資料。
2. 運用監視器、微型攝影機、無人機或其他影像，進行涉及個人辨識、追蹤或隱私之人臉、車牌、生物特徵、人物或行為辨識。
3. 進行治安事件、交通事故、犯罪情資、勤務部署規劃之預測及分析，可能影響人民權益。
4. 建置可存取個資、刑案、報案、影像或其他警政機敏資料之內部 AI 系統。
5. 部署供民眾使用之 AI 聊天機器人或自動化服務。
6. 草擬刑案偵查文書、行政處分、交通舉發、案件通知、陳情處理或其他涉及民眾權利義務之文書。
7. 使用 AI 系統介接本局資訊系統、資料庫或外部服務。
8. 其他經本局評估可能造成重大資安、法遵、隱私或權益風險之應用。

（三）中度風險的使用情境：

協助製作對外文書或內容，與民眾權利義務無關，且不涉及機敏性資料、民眾隱私及個人資料者，通常屬中度風險，例如新聞稿、政策說明、宣導教材及對外統計說明之草擬、編輯或圖表製作。產出應由承辦人員親自檢查、校對及修改，並於對外發布前送請主管審閱。

（四）低度風險的使用情境：

限於內部一般行政輔助，與民眾權利義務、執法判斷無關，且不涉及機敏性資料、民眾隱私及個人資料者，通常屬低度風險，例如內部通知、公開資料摘要或翻譯、一般簡報，以及僅使用公開或非機敏且不含個人資料之內部知識問答。使用人員仍應與原文核對，辦理檢核、核定及適當標示。

五、本局各單位使用 AI 應落實採行以下資料治理與去識別化措施：

- （一）AI 應用所需資料，應依資料來源、資料性質、機密等級、個資屬性、保存期限及使用目的進行分類分級管理。
- （二）涉及個資或機敏資料之應用，應以最小使用為原則，不得過度蒐集

或使用。如確有使用之必要，應先行將姓名、身分證統一編號、出生年月日等得直接或間接識別個人之欄位進行去識別化，使其達到無從識別特定當事人之狀態。並應高度警惕數據重識別之風險，避免因交叉比對，導致已去識別化之資料遭到還原而外洩個資。

(三) 提供外部學術單位、廠商或合作機構使用之資料，應完成必要之去識別化、保密約定、使用範圍限制及成果歸屬約定。

六、本局各單位擬導入、採購、委外開發、試辦、擴充或與外部機關、學術單位、研究機構、廠商合作辦理 AI 應用者，應先填寫附件一檢核表提送本局進行初步評估，規範如下：

(一) 各單位提案時，應初步說明下列事項：

1. 業務需求及預期效益。

2. 使用情境及服務對象。

(1) 預計使用之 AI 技術類型、模型、系統或服務。

(2) 使用資料來源、資料類型及是否涉及個資、機敏資料、公務上應保密資料或民眾權利義務。

(3) 是否需介接本局既有資訊系統、資料庫、監視器影像、一一〇報案資料、交通資料、刑案資料或其他警政資料。

(4) 預估所需算力、儲存、網路、資安及維運資源。

(5) 是否涉及外部廠商、學術單位或其他機關合作。

(6) 預估經費來源、辦理期程及後續維運需求。

(二) 本局 AI 推動小組應就各項提案之業務必要性、技術可行性、資料使用合法性、資安風險、個資保護、算力需求、平台整合可行性、預算合理性及後續維運能力進行研議，並作為本局 AI 應用推動、年度預算編列、科技計畫提報、採購規劃及資源配置之參考。

(三) 涉及下列事項者，應列為重點研議案件，經本局 AI 推動小組討論後，始得進一步辦理試辦、採購、委外或正式導入：

1. 涉及個人資料、機敏資料、公務上應保密資料或犯罪偵查資料。

2. 涉及監視器影像、車牌、臉部、生物特徵、行蹤軌跡或其他可能涉及民眾隱私資料。

3. 涉及一一〇報案、勤務部署、交通執法、犯罪偵查、行政處分或其他可能影響民眾權利義務事項。

4. 涉及跨系統、跨單位、跨機關資料整合。

5. 涉及外部學術單位、研究機構或廠商使用本局資料進行 AI 訓練

、測試、驗證或研究。

6. 涉及新增或大量使用本局 GPU 算力、模型服務、向量資料庫、資料平台或其他共用 AI 資源。

七、本局各單位使用人工智慧須辨識其適用情境、使用風險與限制，預先採取相應措施並尋求資訊科技、法律與倫理等跨單位專業合作，以確保人工智慧的使用符合公眾利益和公共價值。本局各單位使用人工智慧時皆應承擔以下責任：

- (一) 課責性：人工智慧的使用者必須為其所生成之內容負責，包含確保其內容符合事實、且遵循法律與倫理規範。鑑於當今人工智慧所生成之內容仍常出現幻覺與不可接受之錯誤，因此使用人工智慧協助生成文字、圖案或影片時，務必檢視並確認其正確性。
- (二) 安全性：使用人工智慧時應慎選平台與工具並詳細閱讀使用條款，確保符合機密分類規範，以免洩漏國家或政府機密。使用外部或公用人工智慧時，必須假設輸入及產出之內容皆可能被外部機構所記錄。除非透過本局自建或不對外聯網，否則不得使用人工智慧來提問或生成機敏性文書。
- (三) 透明性與可解釋性：使用人工智慧生成文字、語音、圖案或影片等，包含但不限于以上內容時必須清楚標示，就使用人工智慧之情況為適當之揭露，並能明確解釋其提問過程與所生成結果之關聯。如使用人工智慧係以委外採購供應，其契約應要求廠商揭露系統架構、模型部署方式及資料流向，並明定履約期間查核機制。
- (四) 公平且無歧視：使用人工智慧生成文字、圖案或影片時必須確保所生成之內容，對任何族群、種族、性別及社經地位均公平且無歧視。人工智慧模型的訓練樣本可能有取樣偏差而無法代表母體，或訓練時人工標註過程可能有偏見，使得人工智慧所生成之內容帶有偏見歧視，因此對於其所生成內容務必經由人工檢視並於必要時進行修改與確認。
- (五) 隱私與個資保護：對於輸入人工智慧的內容，以及對於人工智慧所生成之內容，應採資料最小化原則，且均應確保不包含民眾之隱私與個人資料，如包含民眾之個人資料應以「去識別化準則」先進行去識別化處理（詳參本府「個資去識別化平臺標準作業流程」）。
- (六) 教育訓練與善用：人工智慧之使用者皆應學習辨識其使用風險與限制，並學習如何運用其優勢以提升工作品質與效率，包含如何對人

工智慧下達貼切的提詞以強化對人工智慧的使用。