「臺北市政府使用人工智慧作業指引」

113 年 9 月 9 日府資創字第 11330102301 號函頒

臺北市政府

中華民國 113 年 9 月 9 日

目錄

壹、目的 1
貳、適用對象 2
參、人工智慧的架構 2
肆、人工智慧之實際應用與使用風險 3
伍、責任與法律遵循 6
陸、使用人工智慧標準作業流程圖及檢核表 9
柒、本作業指引之修正 10
附件一：人工智慧的底層技術 11
附件二：人工智慧的應用開發 13
附件三：臺北市政府使用人工智慧內部控制流程圖 14
附件四：檢核表 15
附件五：使用人工智慧作業常見問題 17
參考文獻 I

臺北市政府使用人工智慧作業指引

壹、目的

近 10 年來，電腦算力的大幅躍進、資料品質與數量的逐步提昇、與演算法的長足進步引領了人工智慧（Artificial Intelligence, AI）爆發性的成展，其已滲入人類生活的每個角落，時時刻刻影響著人類社會。依功能之不同，人工智慧可區分為洞悉數據模式的分析式人工智慧、基於歷史數據預測未來趨勢的預測式人工智慧、以及能生成新數據或新內容的生成式人工智慧。

特別是生成式人工智慧於 2022 年底橫空出世，可以依據使用者輸入的提示詞（prompt）生成文字、語音、圖案、或影片，更是在設計、媒體、軟體、醫藥、教育、司法和理財等領域徹底顛覆了人們的工作型態。公務機關可運用人工智慧協助進行資料分析、城市規劃、交通管理、財政管理與環境資源來增進公務效率，協助醫療保健及教育訓練之執行來提升人民的健康與學養，更可運用生成式人工智慧建置智慧客服來拓展對廣大市民的服務。

然而人工智慧在帶給人們生活便利並提升工作效率的同時，可能也會出現偏見歧視、洩漏機密、危害資訊安全、侵害隱私與個人資料、侵害智慧財產權、發布不恰當言論等等風險。因此，公務機關在使用人工智慧時，應兼顧永續發展與人類自主，並恪守課責性（accountability）、安全性（security）、透明性（transparency）與可解釋性（explainability）、公平且無歧視、隱私與個資保護等倫理原則。此外民法、個人資料保護法及著作權法等相關法律與行政命令之遵守亦成為法律遵循的必要項目，以期最大限度地降低人工智慧的歧視與偏見、降低侵害民眾之個人資料或隱私的風險、同時降低侵害智慧財產權之疑慮。

「臺北市政府使用人工智慧作業指引」（下稱「本作業指引」），乃以行政院所發布之「行政院及所屬機關(構)使用生成式 AI 參考指引」及研議中之「人工智慧基本法草案」為基礎，再參考美國、加拿大、英國、澳洲等國的人工智慧指引先例，進一步說明人工智慧於政府機關的實際應用以及可能衍生之風險，闡釋各機關於使用人工智慧時必須遵循之倫理與法律規範。本作業指引提供使用人工智慧標準作業流程，並臚列各機關使用人工智慧時可能的問題並提供解答，以利各機關於使用人工智慧時有所遵循。除了必須識別前述風險以及諸多法規遵循與倫理考量外，本府於導入人工智慧於內部業務推動或外部民眾服務時，尚須評估風險管理所需廣義成本（包含本作業指引內容或其他相關財務與非財務成本），並蒐集質性或量化證據以研判相對於導入前確實具有品質或效率提昇之效益，同時衡量獲致效益所需廣義成本（財務與非財務）的可行與適切性。

貳、適用對象

一、 本府及所屬各機關，應遵循本作業指引。各機關並得視使用人工智慧之業務需求，參酌本作業指引另訂使用規範或內控管理措施。

二、 本府及所屬各機關就所辦理之採購事項，應要求得標之事業、法人、團體或個人遵循本作業指引，並恪遵各該機關所制定之使用規範或內控管理措施。

三、 本府所屬公立學校及公營事業機構得自訂人工智慧使用指引，並得準用本作業指引。

參、人工智慧的架構

人工智慧的架構包含底層技術、應用開發、與實際運用等三層。其中底層技術包含專家系統（expert system）、機器學習（machine learning）、深度學習（deep learning）及其演算法（algorithm）等，請參閱附件一之詳細說明。中間的應用開發則包含自然語言處理（natural language processing）、機器視覺（machine vision）、生成式人工智慧（generative artificial intelligence）等，請參閱附件二的詳細說明。最上層人工智慧的實際運用為本作業指引的核心，將於第肆條說明人工智慧的實際運用範例及其可能衍生之風險。

一般而言，人工智慧的訓練區分成預訓練（pre-training）與微調（fine-tuning）兩階段。以生成式人工智慧為例，其預訓練通常係由大型科技公司蒐集海量的文本或圖案來訓練語言模型，對預訓練好之語言模型可能選擇閉源或開源或介於兩者間的共享式授權。如果預訓練語言模型的大型科技公司選擇閉源，便僅有其自身或經其授權之機構方可運用其預訓練之語言模型進行微調，再將微調好的語言模型部署至公開網路做商業使用（下稱「外部人工智慧」）。然倘其選擇開源或共享授權，則任何機構皆可在其規範條文內，基於其預訓練之語言模型，以該機構自身所擁有或蒐集所得之專業領域資料進行微調，以讓語言模型符合該機構專業領域的使用需求。

政府機關於使用外部人工智慧時應具有資訊安全之風險意識，宜假設任何輸入之內容皆可能被蒐集利用。因此當政府機關有將機密性或敏感性資料(下稱「機敏性資料」）使用於人工智慧之高度需求時，最好自建不對外連網的人工智慧。

肆、人工智慧之實際應用與使用風險

一、 人工智慧於政府機關的實際應用：

各機關可運用人工智慧並搭配相關資通訊技術之應用範圍相當廣泛，例如（但不限於）下列應用場景：

（一）教育訓練：政府機關可利用人工智慧來提升教育品質與職業訓練，包含但不限於：提升教育行政效率、建置多元教學資源、分析學生學習狀況並推薦客製化輔助教學內容、分析人力市場對職場能力之需求以強化職業訓練、以及草擬教育政策之參考等。

（二）智慧客服：政府機關可使用人工智慧之聊天機器人來提升對市民的服務，包含但不限於：回答市民的常見問題、詢問市民需求並導引至專責機關、協助市民查詢市政相關資訊、及協助市民處理簡單且較為制式之市政事務之參考。

（三）政策行銷：政府機關可運用人工智慧來強化政策行銷與社會溝通，包含但不限於：協助分析輿情、了解民眾需求或痛點、協助依據民眾實際需求傳遞服務民眾之訊息、及協助撰寫政策宣導與民意溝通之說帖的參考。

（四）環境資源：政府機關可運用人工智慧來強化環境保護與自然資源管理，包含但不限於：以人工智慧分析雷達或衛星影像來監測空氣品質、監測水文及水庫狀況、識別並監測環境污染、規劃山林及生態保育、管理天然資源、進行環境保護教育之參考等。

（五）醫療保健：政府機關可利用人工智慧來提升醫療品質與公共健康，包含但不限於：提升醫務管理效能、監控流行病或傳染性疾病之爆發與散播、分析病歷資料與醫療影像、協助新藥開發與藥效預測、提升長期醫療照護之品質、以及草擬醫療政策之參考等。

（六）交通管理：政府機關可運用人工智慧進行智慧交通管理，包含但不限於：監測交通情況、預測堵塞狀況、優化交通流暢度、檢視交通違規、分析交通事故、開發智慧停車系統、以及制定交通政策之參考等。

（七）資料分析：政府機關可運用人工智慧進行資料清理、格式轉換與特徵汲取，從大量資料中挖掘有用資訊並進行分析與可視化，協助釐清問題、分析輿情、進而形成決策之參考。

（八）城市規劃：政府機關可運用人工智慧進行城市規劃並優化決策，包含但不限於：分析市政相關資料、依據市政資料建構多個城市發展藍

圖、對每個市政發展藍圖進行模擬以分析其優缺點並找尋最佳方案、對尋得之最佳方案提出如何落實之具體建議之參考等。

（九）人工智慧助理：上述與其他多元的應用場景中，凡是以 AI 相關技術（如機器學習、自然語言處理等）為基礎所開發的軟體、線上服務、或裝置，足以作為推動市府業務的虛擬助手，皆可廣義稱為人工智慧助理，例如電子郵件處理、日程管理、文件自動生成等。

以上所述人工智慧應用場景僅是舉例，隨著人工智慧越來越成熟與普及，本府對人工智慧之運用必然不以上述場景為限。

二、 人工智慧的使用風險與限制：

市政府使用人工智慧固然可提升行政效能並強化決策能力與品質，惟亦可能出現行政失當甚至違法之風險，茲區分為絕對禁止事項、高度風險的使用態樣、中度風險的使用態樣、以及低度風險的使用態樣，分別說明如下：

（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府同意公開的資料，或未符合法律規範且未經事前同意即使用人工智慧處理或利用民眾之特種個人資料，或未經人工檢視與確認即使用人工智慧直接做成影響民眾權利義務之決定者，通常屬於絕對禁止事項，例如：

1. 當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府同意公開的資料時，除非使用本府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，否則絕對禁止。

2. 未符合法律規範且未經人民之事前同意使用人工智慧對民眾進行人臉、指紋或其他生物特徵之辨識。

3. 未符合法律規範且未經人民之事前同意，使用人工智慧之輸入內容包含民眾之病歷、醫療、基因、性生活、健康檢查或犯罪前科等特種個人資料。

4. 未經人工檢視與確認，使用人工智慧直接做成影響民眾權利義務之行政或法律處分。

5. 未經人工檢視與確認，使用人工智慧直接對人民進行評分或評核。

（二）高度風險的使用情境

當使用人工智慧的情境與民眾之權利義務相關，或涉及民眾之隱私或個人資料，或可能出現偏見歧視時，通常屬於高度風險，例如：

1. 使用人工智慧進行數據分析：尤其可能包含民眾之隱私或個人資訊時，倘未妥善處理可能會有侵害民眾隱私或個人資訊之風險。

2. 部署供民眾使用之聊天機器人：聊天機器人與民眾對談過程中可能出現不當甚至違法言論，諸如具有特定政治傾向、激進或仇恨性言論，或對種族、性別、出生地或居住地、膚色、社經地位等具有偏見或歧視之言論。

3. 使用生成式人工智慧之輸入內容包含民眾之隱私與個人資料：輸入生成式人工智慧之內容，包含但不限於以提詞、情境學習或檢索增強生成方式輸入者，均可能包含民眾之隱私與個人資料。

4. 使用生成式人工智慧撰寫或編輯與民眾權利義務相關之文書：人工智慧所撰寫與民眾權利義務相關之文書可能出現與主旨無關甚至不當或違法用語，亦可能出現對種族、性別、膚色、社經地位等具有偏見或歧視文字。承辦人員務必親自檢查、校對及修改，於對外發布前送請主管審閱，並就使用人工智慧之情況為適當之揭露。

5. 符合法律規範或經人民之事前同意，使用人工智慧對民眾進行人臉、指紋或其他生物特徵之辨識。

（三）中度風險的使用情境

當使用人工智慧的目的係為協助製作對外文書，但與民眾之權利義務無關，且不涉及機敏性資料和民眾之隱私及個人資料時，通常屬於中度風險，例如：

1. 對於未涉機敏性資料且後續須經長官檢閱並批准之文書，使用生成式人工智慧或瀏覽器擴充套件協助撰寫與編輯草稿。

2. 運用生成式人工智慧或瀏覽器擴充套件協助草擬上述例舉各政策領域文件時，應謹記生成式人工智慧並不保證其回答之專業品質，尤其牽涉相對弱勢而無法被妥適反映於訓練文本的利害關係人。因此僅能將其當成研擬政策之請益諮詢或集思廣益輔助，承辦人員與政策制定者仍應自負全部責任。

3. 使用生成式人工智慧或瀏覽器擴充套件撰寫或編輯對外新聞稿、公告、徵才啟事或政策宣導：如同以上與民眾權利義務相關文書，可能出現與主旨無關甚至不當或違法用語，亦可能出現對種族、性

別、膚色、社經地位等具有偏見或歧視文字。承辦人員務必親自檢查、校對及修改確認，並於對外發布前送請主管審閱。

（四）低度風險的使用情境

當使用人工智慧的情境限於市政府內部作業，無關民眾之權利義務，亦不涉及民眾之隱私及個人資料時，通常屬於低度風險，例如：

1. 使用生成式人工智慧或瀏覽器擴充套件協助撰寫機關內部之電子郵件，或輸入資料及輸出成果均僅限內部人員檢視及使用之情形，例如：通知員工於颱風期間注意安全、本府單位對內部人員提供之智慧客服。

2. 面對龐大的公開資料或文件，可運用生成式人工智慧或瀏覽器擴充套件協助擷取重點或進行摘要。惟所擷取之重點或摘要可能簡化、忽略甚至誤解原文（尤其領域相關專業術語），承辦人員仍應與原文對照以確保內容之正確性。

3. 面對繁複的外文公開資料或文件，可運用生成式人工智慧協助翻譯成中文並擷取重點。惟生成式人工智慧之翻譯可能不夠精準，所擷取之重點或摘要可能簡化、忽略甚至誤解原文（尤其領域相關專業術語），承辦人員仍應與原文對照以確保翻譯之正確性。

伍、責任與法律遵循

一、 市政府各機關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、法律與倫理等跨部門專業合作，以確保人工智慧的使用符合公眾利益和公共價值。各機關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。除非透過本府自建或不對外聯網，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…包含但不限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連。

（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，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。

二、 市政府各機關使用人工智慧時，更須遵循各種法律和行政命令。除應依行政程序法之規範落實依法行政及依公務員服務法之規範執行公務外，尚應遵守之法律規範包含但不限於：

（一）民法對隱私的保護：民眾的隱私屬於人格權之一環，依據民法第 18 條之規定，當民眾之人格權受侵害時，得請求法院除去其侵害；有受侵害之虞時，得請求防止之；當法律有特別規定時，並得請求損害賠償或慰撫金。民法第 195 條第 1 項則直接載明對民眾隱私之保護，當民眾之隱私被侵害時，縱使沒有財產上之損害，亦得請求賠償相當之金額。故而使用人工智慧時必須檢查是否涉及民眾之隱私，例如：對生成式人工智慧所輸入之提詞即不得包含民眾之隱私，人工智慧所生成之結果倘包含民眾隱私，應先將其刪除後方得使用之。

（二）個人資料保護法：依據個人資料保護法（下稱「個資法」）第 2 條第 1 款之規定，舉凡「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」均屬個人資料。使用人工智慧時如有涉及民眾個人資料之蒐集、處理及利用，必須嚴守個資法之規範。例如：對生成式人工智慧輸入提詞時，倘提詞中包含民眾的個人資料，即屬對個人資料的利用。此時應盡可能將民眾之個

人資料去識別化，倘無法去識別化，則應再檢視此時對民眾個人資料之利用是否係為執行法定職務所必要，並與蒐集之特定目的相符。若否，則除符合以下情事之一者外，均不得利用：（一）法律明文規定；（二）為維護國家安全或增進公共利益所必要；（三）為免除當事人之生命、身體、自由或財產上之危險；（四）為防止他人權益之重大危害；（五）公務機關或學術研究機構基於公共利益為統計或學術研究而有必要，且資料經過提供者處理後或經蒐集者依其揭露方式無從識別特定之當事人；（六）有利於當事人權益；或（七）經當事人同意。此外對於人工智慧所生成之內容，亦須檢視其是否包含民眾之個人資料，如是，同樣須踐行前述去識別化，或確認對民眾個人資料之利用乃於執行法定職務必要範圍內為之，並與蒐集之特定目的相符。

(三)著作權法：著作權包含著作人格權與著作財產權。由於人工智慧之訓練過程必須使用海量資料作為訓練文本，且大多數訓練文本都受著作權之保護，因此國際上對於人工智慧侵權的指控不斷，衍生許多糾紛。因此在使用人工智慧生成文字、語音、圖案或影片時，應與資料庫及網路上之既有內容相比對。當發現人工智慧所生成之文字、語音、圖案或影片與既有內容，於著作表達形式上完全相同或實質近似時，應以人工大幅修改或重新演算輸出，方得對外公開或使用，除非個案上明確符合公務利用之合理使用則不受此限，以免徒生侵害著作人格權與著作財產權之爭議。

(四)資通安全管理法：為積極推動國家資通安全政策並加速建構國家資通安全環境，我國制定有資通安全管理法，以保障國家安全並維護社會公共利益。依據該法之規範，公務機關應符合其所屬資通安全責任等級之要求，並考量其所保有或處理之資訊種類、數量、性質、資通系統之規模與性質等條件，訂定、修正及實施資通安全維護計畫。基此，為強化所屬各機關（構）及受監督行政法人（各機關）資通安全管理，建立安全及可信賴之電子化政府，本府定有「臺北市政府資通安全管理規定」，以確保資料、系統、設備、網路安全及人員安全，降低因人為疏失、蓄意或天然災害等導致之資通訊資產遭竊、不當使用、洩漏、竄改或破壞等風險。

陸、使用人工智慧標準作業流程圖及檢核表

一、 於本府公務使用人工智慧（含所有應用場景，請一併參照本作業指引附件三之臺北市政府使用人工智慧內部控制流程圖），應遵循以下事項：

（一）先確定使用人工智慧之目的範圍及預期成效，並掌握其可能風險、限制與適用情境，對於前述「絕對禁止事項」(詳第肆條第 2 項第 1 款)不得使用人工智慧。

（二）申請外部人工智慧帳號時，應使用本府員工電子郵件帳號。

（三）仔細閱讀並了解人工智慧的使用條款，若相關條款使用比例低，或嵌入可能妨礙公務目標施行之爭議條件，宜避免使用，必要時得徵詢本府資料治理委員會意見。

（四）盡可能選擇拒絕或退出（opt out）人工智慧蒐集並記錄所輸入資料，然若欲藉累積性資料之互動提升後續服務品質，且所輸入資料不涉及任何職務責任之違反，並對民眾法定權利之保護者無損害風險者，得例外處理之。

（五）對外發布時，應揭露所使用人工智慧之廠商名稱、模型名稱及其版本。

（六）如牽涉民眾隱私、個資或本府機敏性資料，應由本府自建或訓練調校不對外聯網的人工智慧模型，並詳參本府「個資去識別化平臺標準作業流程」。

（七）確保人工智慧之訓練模型、訓練資料、與產出成果符合課責性、安全性、透明性與可解釋性、公平且無歧視等倫理原則，並遵循個資、著作權、與各專業領域相關法規。

（八）應逐一確認本作業指引附件四「檢核表」之項目是否均符合規定。

二、 於本府公務使用生成式人工智慧或瀏覽器擴充套件（除前開使用人工智慧之標準作業流程外，更包含下列使用流程），應遵循以下事項：

（一）務必檢查提示詞或檢索字詞是否涉及不得或不宜公開之內容，例如：政府機密、民眾之隱私與個人資料等。

（二）針對人工智慧或瀏覽器擴充套件所生成內容，務必檢查其所產出之內容是否有違反法規、侵害民眾隱私或個人資料、或侵害著作權之虞。

（三）應逐一確認本作業指引附件四「檢核表」之項目是否均符合規定。

（四）應註明其內容係由何種生成式人工智慧或瀏覽器擴充套件所草擬並經承辦人員親自檢查、校對及修改，並於對外發布前送請主管審閱。

（五）本府於使用外部人工智慧時應具有資訊安全之風險意識，宜假設任何輸入之內容皆可能被蒐集利用。因此當本府有將機敏性資料使用於生成式人工智慧之高度需求時，最好自建不對外連網的大型語言模型，其步驟包含：挑選開源或共享授權的預訓練語言模型、以自身所擁有或蒐集之專業資料進行微調、經測試並驗證其正確性後、部署至機構內部僅供機構內使用。

柒、本作業指引之修正

本作業指引將依中央或臺北市政府法規之制定或修正、或當人工智慧的技術有大幅躍進時，定期或不定期進行修正。

附件一：人工智慧的底層技術

一、專家系統（Expert System）：

專家系統乃發展較早的人工智慧技術，其依賴儲存特定領域相關知識與規則的知識庫（knowledge base）來模擬特定領域專家的推理與判斷能力，以協助使用者於特定領域進行決策分析或解決問題。

專家系統比較適用於能制定出「若 A 則 B」之明確規則的領域，例如於疾病診斷上「若醫事檢驗結果發現某某徵狀」，則「病患可能罹患某種疾病」等。專家系統之效能通常與知識庫（包含特定領域之相關知識與規則）之豐富度與完整性成正比，須仰賴大量專家之投入以提供特定領域相關知識並制定規則，常被應用於都市規劃、醫療診斷、污染防治等。

二、機器學習（Machine Learning）：

機器學習是人工智慧技術的重要分支，不需如專家系統般地先就特定領域建立規則，而是讓分類或預測等演算法透過大量資料之學習以自動找出資料的規律與模式，從而建構人工智慧模型。機器學習被廣泛地運用於文字識別、語音辨識、及圖形辨識等領域。一般而言資料量越多且品質越好，所訓練出人工智慧模型的效能越佳。機器學習的步驟包含：(1) 準備訓練資料：包含資料之蒐集、清洗、與預處理；(2) 訓練模型：將前述預備好之訓練資料輸入演算法中以調整演算法之參數，令其盡可能符合訓練資料之規律與模式；(3) 測試與評估模型：測試並評估訓練完成模型之效能，必要時必須重複訓練以調整參數，如重複訓練效果依然不佳時甚至必須以其他模型再行訓練、測試與評估；(4) 部署人工智慧模型：將前述訓練、測試與評估完成之模型部署至實際運用。

依據訓練方式之不同，機器學習可區分為監督式學習（supervised learning）、非監督式學習（non-supervised learning）和強化學習（reinforcement learning）等不同訓練方式。監督式學習必須先由人工對訓練資料進行標註，之後將訓練資料輸入演算法中讓電腦程式自動調整演算法之參數，使其輸出結果符合人工標註之標籤。非監督式學習則不須以人工對訓練資料進行標註，而是透過降維或分群等演算法讓電腦程式自行發覺資料的規律與模式。強化學習則是透過評分機制與獎勵措施的制定，讓人工智慧進行自我評估並朝獲取最大獎勵的方向進行學習。

三、深度學習（Deep Learning）：

深度學習是機器學習的一個分支，通過多層的人工神經網路來學習大量之文本、語音、圖片或影像。人工神經網路的每一層皆由多數個彼此連結的神經元所組成，用以處理資訊並相互傳遞訊號。每個神經元均可接收前一層人工神經網路某神經元的輸入，以激勵函數處理後再輸出至人工神經網路的下一層。當使用大

量學習資料進行深度學習時，學習資料會調整神經元之間的連接，從而自動萃取出訓練資料的特徵。

與傳統機器學習的主要差異是，深度學習不需人工給定特徵，而是藉由人工神經網路之運作，在人工智慧的學習過程中自動挖掘出資料的特徵。深度學習的主要應用包含自然語言處理、影像辨識與處理、語音辨識與合成、與生成式人工智慧等。

附件二：人工智慧的應用開發

一、自然語言處理：

自然語言處理是人工智慧技術與語言學之跨領域整合，旨在辨識與理解人類的語音或文字，並能以語音或文字做成相對應之回應。依據語音或文字之輸入與輸出的排列組合，自然語言處理之實際運用包含：(1) 語音轉換成語音：如語音聊天機器人或不同語言間的同步口譯；(2) 語音轉換成文字：如自動繕打會議逐字稿或辨識外國語言再翻譯成中文字幕；(3) 文字轉換成語音：例如將書籍內容朗讀出來給讀者聆聽的有聲書；(4)文字轉換成文字：如文本翻譯、文本內容摘要、文本語意分析或情感分析、資訊檢索、及文字聊天機器人等等。

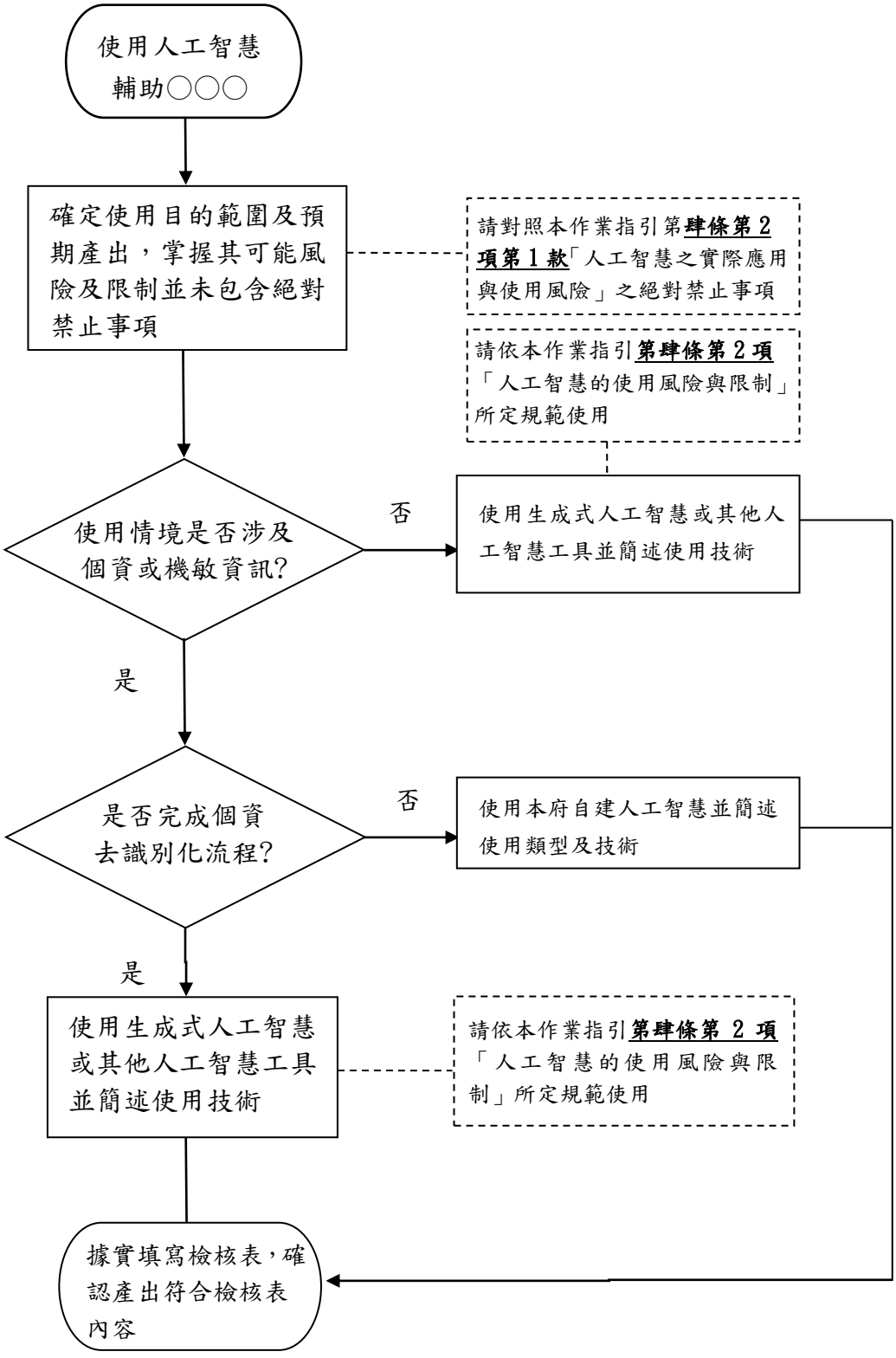
二、機器視覺：

機器視覺是影像辨識與處理的技術。為以電腦科技與人工智慧來實現人眼之視覺效果，機器視覺必須先以照相機、錄影機或其他感測裝置擷取待判別物體之影像，再進行影像分析以識別物體之大小、形狀、遠近、花紋、色澤及明暗等特徵，據以判斷該物體之種類與狀態以執行進一步決策。無人車是其中一項重要的應用，其以光達、雷達及攝影機等裝置擷取前方路況，當擷取到路邊有物體之影像時可藉由分析其大小、形狀、遠近、色澤等特徵分辨該物體是行人、消防栓還是變電箱，據以做出是否緊急煞車或減速或可安全通行之決策。

三、生成式人工智慧：

生成式人工智慧乃基於富含數千億個參數的大型語言模型，在使用者下達「提詞」後，可針對使用者的提問依據文字機率分布以文字接龍方式輸出，來回應使用者之提問。生成式人工智慧對提詞的理解能力與因應提問的文字生成能力，皆與其語言模型的參數數量呈現正相關。生成式人工智慧的訓練區分成預訓練與微調兩階段，其中預訓練係蒐集海量的文本來訓練語言模型，微調則是以較少量的專業領域資料對預訓練模型進行調教。

附件三：臺北市政府使用人工智慧內部控制流程圖



附件四：檢核表

臺北市政府使用人工智慧檢核表

一	承辦機關 （主協辦）	
二	公務項目	標題： 請簡述：
三	運用範圍	（一）網域(AI引擎)（可複選）： □外部網域 □本府自建(地端)□不對外聯網之地端機房 □相同資安防護等級之自控機房□其他 請簡述： （二）使用類型（可複選）： □分析式人工智慧 □預測式人工智慧 □生成式人工智慧 □其他 請簡述：人工智慧的廠商名稱、模型名稱與版本 （三）運用方式（可複選）(參考本作業指引第肆條第1項)： □政策規劃 □資料分析 □教育訓練 □智慧客服 □環境監測 □文稿產出 □其他 請簡述： （四）輸入或訓練資料類型（可複選）： □非機敏資訊 □內部去識別化資料 □個資或機敏資訊 □其他 請簡述：

		（五）輸入或訓練資料來源（可複選）： □本局處內部　□本府其他局處 □府外來源　□其他
		請簡述：
		（六）運用範圍是否包含本作業指引所列絕對禁止事項 **（參考本作業指引第肆條第 2 項第 1 款）：** □是　□否　□其他
		請簡述：
		（七）產出成果是否直接影響民眾權利義務： □是　□否　□其他
		請簡述：
		（八）本人(機關)充分瞭解使用人工智慧之責任及法遵事項 **（參考本作業指引第伍條）：** □是　□否
四	**風險等級**	□低　□中　□高（參考本作業指引第肆條第 2 項第 1 款）
五	**其他 補充說明**	
六	**填表人 單位職稱/姓名**	單位職稱： 姓名：
備註	一、 分析式人工智慧：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。 二、 預測式人工智慧：可基於歷史數據來預測未來可能發生的事件、趨勢、或結果。 三、 生成式人工智慧：可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。	

附件五：使用人工智慧作業常見問題

一、 可否使用個人的電子郵件信箱來註冊公務所使用人工智慧的帳戶？

答：不可以。

如係為市政府公務所使用，應以本府員工電子郵件帳號來註冊人工智慧的帳戶，或使用資訊局提供之公務帳號，以確保使用記錄被妥善留存，並可視需要移轉至市政府之資料庫。

二、 可否直接採信人工智慧評價或判斷民眾相關權利義務？

答：不可以。

人工智慧對民眾行為之判斷或評價僅能作為參考，必須經由主管機關之人為判斷裁量才能對民眾相關權利義務予以評判。

三、 可否使用外部生成式人工智慧來研擬機密性文書？

答：不可以。

考量輸入外部生成式人工智慧的內容及其產出的內容可能會被外部機構所記錄，衍生機密資料外流之疑慮，因此禁止輸入及輸出機密文書。如有機密文書使用需求，請依相關規定使用市政府自建或不對外聯網之人工智慧。

四、 可否使用外部生成式人工智慧來研擬行政處分文書?

答：不可以。

考量輸入外部生成式人工智慧的內容及其產出的內容可能會被外部機構所記錄，衍生機敏資料外流之疑慮，由於行政處分文書通常涉及民眾的個人資料與權利義務事項，如有必要使用之需求，請循相關規定使用市政府自建或不對外聯網之人工智慧。

五、 可否使用外部生成式人工智慧來撰寫電腦程式？

答：依該電腦程式使用目的及資料風險等級而定。

如撰寫之電腦程式涉及機敏事項時，不得使用外部生成式人工智慧進行撰寫；如未涉及機敏事項，原則上可使用外部生成式人工智慧來協助撰寫電腦程式，惟應以近似度比對軟體進行近似度比對。當發現該人工智慧所撰寫之電腦程式與既有之開源軟體或共享軟體之程式碼相同或實質近似時，應遵循個別開源授權條款或共享授權條款；當發現該人工智慧所撰寫之電腦程式與既有之第三方軟體相同或實質近似時，應以人工大幅修改或重新演算輸出，方得對外公開或使用。

六、 可否使用外部生成式人工智慧來研擬本府內部報告或電子郵件?

答：可以，惟須經檢視確認後始得發布或寄出。

使用外部生成式人工智慧來研擬市府內部報告或電子郵件時應確認提詞中不應包含政府機密資訊、民眾隱私及個人資料，如涉及民眾個人資料應依「臺北市政府個資去識別化平臺標準作業流程」之規定完成去識別化作業。針對人工智慧產出之內容，承辦人員應自行檢視及修改，以確保產出內容正確、無偏見或歧視且未侵害他人著作權。另於擬發布之內部報告或電子郵件中，應揭示其內容係運用何種人工智慧所產生，並經承辦人員及主管檢視確認後始得發布或寄出。

七、 可否使用外部生成式人工智慧來研擬市政府對外文書，例如新聞稿、社交媒體發文、或活動文宣品？

答：可以，惟須經檢視確認後始得發布或寄出。

使用外部或公用生成式人工智慧來研擬市府對外文書時應確認提詞內容不包含任何政府機密資料、民眾隱私及個人資料，如涉及民眾個人資料應依「臺北市政府個資去識別化平臺標準作業流程」之規定完成去識別化作業。針對產出之內容，應自行檢視及修改，以確保產出內容正確、無偏見或歧視且未侵害他人著作權。另於擬發布之對外文書中，應揭示其內容係運用何種人工智慧所產生，並經承辦人員及主管檢視確認後始得發布或寄出。

八、 制定政策時可否使用生成式人工智慧？

答：可以，但僅供輔助及參考使用。

承辦人員於制定政策時可使用外部生成式人工智慧來進行腦力激盪以收集思廣益之效，但仍應留意提詞內容不得包含任何政府機密資料、民眾隱私及個人資訊，如涉及機敏資訊及個資疑慮時，應先完成去識別化作業。另於使用外部生成式人工智慧協助制定政策時，應留意其偶會出現幻覺或過時資訊之侷限性，對其所提供事實、學理、法規及證據不可盡信，應確實進行查證。

九、 使用生成式人工智慧來生成文字或圖畫，是否可能侵害他人的著作權？

答：有可能，避免使用具著作權之特定人物、作者、藝術家、畫作、雕刻、或建築等著作為提示詞。

由於外部生成式人工智慧係經海量資料訓練而成，惟訓練資料未必擁有合法授權，仍有被控侵權之風險。因此當使用外部生成式人工智慧來生成文字或圖畫時，仍應進一步檢索或透過比對系統進行比對及校正工作，避免侵害他人之著作權。如著作表達形式上完全相同或實質近似時，應以人工大幅修改或重新演算輸出，方得對外公開或使用，

除非個案上明確符合公務利用之合理使用則不受此限，以免徒生侵害著作人格權與著作財產權之爭議。

十、 前述高度風險的使用情境中所述「符合法律規範或經人民之事前同意，使用人工智慧對民眾進行人臉、指紋或其他生物特徵之辨識」，包含哪些例子？

答：經人民之事前同意者，例如針對失智或經禁治產宣告之無行為能力人，其親屬至警察局或社會局登錄其指紋等生物特徵。

參考文獻

1. 科技部（現為國科會），人工智慧科研發展指引，科技部（現為國科會）108 年 9 月版，網址：https://www.nstc.gov.tw/folksonomy/detail/dbf8da09-22be-4ef1-8294-8832fc6e8a26?l=ch。
2. 行政院，行政院及所屬機關（構）使用生成式 AI 參考指引（草案），網址：https://www.nstc.gov.tw/folksonomy/detail/f9242c02-6c3b-4289-8e38-b8daa7ab8a75?l=ch。
3. Digital Transformation Agency of Australian Government, Interim guidance on government use of public generative AI tools - November 2023, Australian Government Architecture (Nov. 22, 2023), https://architecture.digital.gov.au/guidance-generative-ai.
4. Government of Canada, Guide on the Use of Generative AI, https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/guide-use-generative-ai.html.
5. Cabinet Office of UK, Guidance to Civil Servants on Use of Generative AI, GOV.UK (Jan. 29, 2024), https://www.gov.uk/government/publications/guidance-to-civil-servants-on-use-of-generative-ai/guidance-to-civil-servants-on-use-of-generative-ai.
6. City of San Jose, Generative AI Guidelines, City of San Jose (Sept. 23, 2023), https://www.sanjoseca.gov/home/showpublisheddocument/100095/638314083307070000.
7. Fui-Hoon Nah, F., Zheng, R., Cai, J., Siau, K., & Chen, L. (2023). Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. Journal of Information Technology Case and Application Research, 25(3), 277-304.
8. Luckett, J. (2023). Regulating Generative AI: A Pathway to Ethical and Responsible Implementation. Journal of Computing Sciences in Colleges, 39(3), 47-6

9. Bright, J., Enock, F. E., Esnaashari, S., Francis, J., Hashem, Y., & Morgan, D. (2024). Generative AI is already widespread in the public sector. arXiv preprint arXiv:2401.01291.
10. Cantens, T. (2024). How will the state think with ChatGPT? The challenges of generative artificial intelligence for public administrations. AI & SOCIETY, 1-12.