連江縣政府人工智慧應用風險評估檢核表

填寫目的：參酌人工智慧基本法規定及人工智慧風險分類框架模式，評估並掌握本府人工智慧系統與服務之潛在風險。

填寫範圍：政府機關自行開發、委外建置或採購之各類人工智慧系統與服務。

填寫說明：本表由人工智慧系統與服務管理人或機關指定人員填寫，作為符合人工智慧基本法第十九條進行風險評估與訂定內控管理措施之基礎。

1、 應用情境盤點

本表格目的為掌握人工智慧系統與服務之基本資訊。

檢核項目	填寫內容/評估結果	依據來源	備註/參考文件
1.1 **系統/服務名稱**	[請填寫]	人工智慧風險分類框架(附錄1)	系統清冊
1.2 **系統/服務具體用途與服務對象**	**【具體用途】** [請填寫] **【服務對象】** □ 政府(跨機關)內部使用 □ 對外公開服務	人工智慧風險分類框架(附錄1)	系統清冊
1.3 核心人工智慧技術與開發方式			
1.3.1 **核心人工智慧技術與開發方式** 大型語言模型（Large Language Model, LLM）	□自行建置（含委外建置） □採購 □未使用	人工智慧風險分類框架(附錄1)	採購或系統文件

1.3.2 核心人工智慧技術與開發方式 生成式 AI 使用(如業務輔助、電腦視覺等)	□自行建置 （含委外建置） □採購 □未使用		
1.4 辨識利害關係人			
負責開發人工智慧系統/服務	□自行開發 □委外開發 □僅進行採購(類 SaaS 服務)	人工智慧風險分類框架(附錄 1)	業務作業說明
負責佈署人工智慧系統/服務	□自行佈署 □委外佈署 □僅進行採購(類 SaaS 服務)		
可能因人工智慧系統/服務影響之利害關係人 *可複選	□民眾 □本機關公務員 □跨機關公務員 □其他：_____		

2、 核心風險識別與危害評估

參考人工智慧風險分類框架，檢視並確認系統與服務是否存在風險。若檢核項目2.1、2.2、2.3任一項勾選「是」，請接續進行檢核項目2.4嚴重程度判定。

檢核項目	評估結果	情境簡述	依據來源
2.0風險類型0：資安及個資問題 本人工智慧系統/服務、訓練資料、輸入指令、輸出內容，是否有可能因其業務流程、使用情境或設計	□是 □否	[若評估結果為是，請簡述]	

邏輯，而有違反資通安全或個資保護之疑慮？			
2.1 風險類型一：人工智慧系統/服務本身之技術設計缺陷 本人工智慧系統/服務，是否有可能因其業務流程、使用情境或設計邏輯，而發生演算法偏見、產生錯誤誤導訊息、缺乏透明度安全漏洞等情況，導致對特定群體不公平或決策錯誤？	□是 □否	[若評估結果為是，請簡述]	人工智慧風險分類框架（二、識別風險）
2.2. 風險類型二：部署後操作及人機互動問題			
2.2.1 本人工智慧系統/服務，是否有可能因其業務流程、使用情境或設計邏輯，導致公務員過度依賴而喪失獨立判斷能力？	□是 □否	[若評估結果為是，請簡述]	人工智慧風險分類框架（二、識別風險）
2.2.2 本人工智慧系統/服務，是否有可能因其業務流程、使用情境或設計邏輯，導致人工智慧系統/服務遭濫用以生成違法內容、深偽詐欺進行網路攻擊?	□是 □否	[若評估結果為是，請簡述]	

2.2.2 本人工智慧系統/服務，是否有可能因其業務流程、使用情境或設計邏輯，導致人工智慧系統/服務遭濫用以生成違法內容、深偽詐欺進行網路攻擊?	□是 □否	[若評估結果為是，請簡述]	
2.2.3 本人工智慧系統/服務，是否有可能因其業務流程、使用情境或設計邏輯，導致人工智慧系統/服務遭利用於網路攻擊？	□是 □否	[若評估結果為是，請簡述]	
2.3 風險類型三：社會結構與環境衝擊			
2.3.1 本人工智慧系統/服務，是否有可能因其業務流程、使用情境或設計邏輯之應用，可能加劇社會不平等之疑慮或風險？	□是 □否	[若評估結果為是，請簡述]	人工智慧風險分類框架（二、識別風險）
2.3.2 本人工智慧系統/服務，是否有可能因其業務流程、使用情境或設計邏輯之應用，可能耗費能源造成環境傷害？	□是 □否	[若評估結果為是，請簡述]	
2.3.3 本人工智慧系統/服務，是否有可能因其業務流程、使用情境或設計邏輯之應	□是 □否	[若評估結果為是，請簡述]	

用，可能遭利用於認知作戰，影響公眾信任？			
2.4 嚴重程度判定 若檢核項目2.1、2.2、2.3任一項勾選「是」，請接續進行嚴重程度判定。			
嚴重程度 （依風險發生時之危害程度判定）		**評估結果**	**依據來源**
2.4.1 【非常嚴重】 符合下列情形之一者屬之： 1. 上開風險發生，可能對機關之營運、資產或信譽等方面產生非常嚴重或災難性之影響。 2. 上開風險發生，可能影響他人合法權益或機關執行業務之公正性及正當性，並使機關所屬人員負刑事責任。 3. 上開風險發生，可能導致特定群體（如特定性別、年齡、身心障礙或弱勢族群）在政府資源分配、資格審查或行政決策上遭受嚴重不公平對待。 4. 上開風險發生，可能違反兒少最佳利益原則，或對其身心發展造成嚴重負面影響。 5. 上開風險發生，可能引發社會恐慌，或造成對人類基本權利、生命安全、財產保障、社會秩序或政府公信力之嚴重危害。		□是 □否	參考人工智慧基本法、「資通安全責任等級分級辦法」之「附表九、資通系統防護需求分級原則」
2.4.2 【嚴重】 符合下列情形之一者屬之： 1. 上開風險發生，可能對機關之營運、資產或信譽等方面產生嚴重		□是 □否	參考人工智慧基本法、「資通安全責任等級分級辦法」

之影響。 2. 上開風險發生，可能影響他人合法權益或機關執行業務之公正性及正當性，並使機關或其所屬人員受行政罰、懲戒或懲處。 3. 上開風險發生，可能導致特定群體（如特定性別、年齡、身心障礙或弱勢族群）在政府資源分配、資格審查或行政決策上遭受不公平對待。 4. 上開風險發生，可能違反兒少最佳利益原則，或對其身心發展造成負面影響。 5. 上開風險發生，可能引發社會恐慌，或造成對人類基本權利、生命安全、財產保障、社會秩序或政府公信力之危害。		之「附表九、資通系統防護需求分級原則」
2.4.3 **【一般】** 符合下列情形之一者屬之： 1. 上開風險發生，可能對機關之營運、資產或信譽等方面產生有限之影響。 2. 上開風險發生，可能發生違反相關規範之情形。	□是 □否	參考「資通安全責任等級分級辦法」之「附表九、資通系統防護需求分級原則」

2.5 風險發生機率 若檢核項目2.4.1、2.4.2、2.4.3任一項勾選「是」，請接續進行**風險發生機率**判定。		
發生機率	□高 □中 □低	

3、 **人工智慧內控管理與應對機制（Internal Controls & Response）**

依檢核項目2.0、2.4、2.5判定結果：

1. 若2.0勾選「是」，該項風險請橋接至機關既有之個資與資安管理機制處理，並提供佐證文件(機關資安維護計畫/個資風險評估紀錄)。
2. **綜合判定結果：**

發生機率 嚴重程度	高	中	低
非常嚴重	甲	甲	乙
嚴重	甲	乙	丙
一般	乙	丙	丙

(1) 若2.4及2.5勾選後落在甲類者：本段所列內容為必辦項目。

(2) 若2.4及2.5勾選後落在乙類者：本段所列內容為必辦項目，惟落實順序次於甲類。

(3) 若2.4及2.5勾選後落在丙類者：本段3.1及3.2所列內容為必辦項目。

(4) 若檢核項目2.1至2.3均為「否」，或2.4.1至2.4.3均為「否」為**無顯著風險**。

3. 必辦項目有不適用情形者，例如系統與服務性質確實無涉決策(3.1)或內容產出(3.2)者，得勾選「不適用」並敘明理由。

檢核項目	落實結果	參考來源
3.1人類監督保留：針對人工智慧提供之決策，作業流程中是否設有適當之人	□是 □否	人工智慧基本法第4條第2款

為介入與最終覆核機制，避免完全自動裁決？	□不適用	
3.2 透明標示：系統與服務之產出是否進行適當資訊揭露(如標註「由人工智慧生成」)?	□是 □否 □不適用	人工智慧基本法第5條
3.3 明確警語：若屬高風險，是否依法明確標示注意事項或警語？	□是 □否 □不適用	人工智慧基本法第5條
3.4 建立救濟機制：針對受人工智慧輔助決策影響之民眾，是否提供明確之申訴管道、決策解釋與權益救濟機制？	□是 □否 □不適用	人工智慧基本法第17條
3.5 強化防護：是否於採購契約或系統與服務規格中，加重對供應商之資安、資料治理與防範歧視偏見之要求？	□是 □否 □不適用	人工智慧基本法第4條第4款、第6款

系統與服務管理人：_____

單位主管：_____