公共人資 24 題線上課程

第 1 題學生講義

115 年高考三級第 3 題

AI 協助公務決策之風險

一、題目

115 年公務人員高等考試三級考試｜人事行政｜公共人力資源管理

三、公務機關的管理者可以透過人工智慧協助決策，其有那些風險或缺點需要注意？（25 分）

二、本題思考過程

審題應由題目文字逐步形成答案範圍與架構。本題可依「作答情境—AI 角色—命題要求—作答範圍—架構判斷」的順序思考。

（一）確認作答情境：公務機關的管理決策

題目以「公務機關的管理者」為作答主體，AI 運用發生於政府管理活動。答案除處理效率與正確，也須納入依法行政、公平對待、程序保障、人民權益與行政課責。

（二）確認 AI 角色：協助管理者形成決策

AI 可能參與資料整理、趨勢預測、方案生成、風險評估或選項排序，會改變管理者取得資訊、比較方案及形成判斷的方式。

（三）確認命題要求：分析風險或缺點

作答主體應提出風險類型，並交代風險形成原因及其對決策品質、人民權益或行政責任的影響。AI 的功能可在前言簡要交代。

（四）確認作答範圍：技術、使用與公共治理

完整答案應涵蓋資料、模型與輸出的技術風險，管理者過度依賴與能力弱化的使用風險，以及公平、透明、隱私、裁量、救濟與問責的治理風險。

（五）形成審題結論

應考人須以公務機關管理者立場，分析 AI 協助決策的風險如何形成、造成何種影響，以及機關與管理者應注意的控制重點。

審題結論

本題須同時回答技術、人員使用及公共治理風險，並連結決策正確性、公平性、透明性、人民權益與行政責任。

三、解題破題與 PHRM 定位

（一）本題核心

本題核心在於建立一套有邏輯的風險分類，說明 AI 為何可能造成錯誤或不當影響，並連結公務決策的正確性、公平性、透明性、人民權益及行政責任。

（二）公共人力資源管理定位

AI 進入人力規劃、招募甄選、考績評量、陞遷、訓練發展及人才配置後，可能影響功績原則、平等機會、程序公平、員工信任與主管判斷責任。作答須至少帶入一項具體人事決策情境，使答案維持公共人力資源管理的學科定位。

（三）主要作答方向

- 建立清楚的風險分類，使資料、模型、人員及治理問題形成一致主線。
- 每項風險均說明形成原因與可能後果，呈現因果關係。
- 連結依法行政、公平、透明、隱私、裁量、程序參與及救濟。
- 強調管理者保有專業判斷與人工覆核，機關建立權責、內控及持續監測。

四、架構判斷：本題可形成兩個完整解題方向

1. 題目未指定分類標準：題目僅要求分析風險或缺點，未限定依技術類型、決策階段或公共價值分類。

2. 存在兩個實質不同的分析軸：第一個方向依風險發生於 AI 決策流程中的位置分析；第二個方向依 AI 可能損害的公務決策價值分析。

3. 兩個方向均能完整回答題目：兩者都可涵蓋 AI 本身限制、管理者使用風險及公共部門治理問題，並各自形成完整前言、內容與結論。

4. 兩個方向各有獨立分類邏輯：流程風險法回答「風險發生在哪個環節」；公共價值治理法回答「AI 可能損害哪些公務決策價值」。

> **判斷結果**
> 本題適合提供兩套完整解法；考場選擇其中一套作答。

兩個解題方向

解題路徑	核心問題	主要分析軸與特色
版本一：AI 決策流程風險法	風險發生在哪一個決策環節？	資料輸入→模型運算→結果輸出→人員使用→責任治理。架構具流程性，便於完整盤點。
版本二：公共價值治理法	AI 可能損害哪些公務決策價值？	正確有效→公平不歧視→透明可解釋→隱私安全→人類自主→問責救濟。公共部門特性較鮮明。

五、兩套解法的共同內容

- AI 輸出品質受資料及模型限制，可能產生偏誤、錯誤或不合情境的建議。
- 管理者可能產生自動化偏誤，使輔助工具實質支配決策。
- 公務決策須維護公平、透明、隱私、程序保障與行政責任。
- 人事決策應兼顧功績原則、平等機會與員工信任。
- 機關、承辦人、主管與廠商之間須建立責任鏈、監督及持續檢驗機制。

六、本題作答不可遺漏的重點

下列內容屬於答案中應實際處理的概念、構面或責任要求。每項均應進一步說明形成原因、可能影響或管理要求。

類別	必要作答重點	作答時應處理的內容
共同必要重點	資料或模型限制	資料品質、代表性、時效、模型設定及情境理解不足，會影響 AI 判斷基礎。
	錯誤或偏誤輸出	錯誤預測、生成式 AI 幻覺、不當推論或歷史偏誤，可能進入決策並擴大行政影響。
	自動化偏誤	管理者過度信賴系統結果，忽略反證、現場資訊、個案差異及專業判斷。
	公平、透明與隱私	公務決策須維護功績原則、平等機會、理由說明、程序保障與資料治理。
	人工覆核與行政責任	AI 維持輔助定位，法定職權者仍須查證、說理並負最終決策責任。
流程路徑必要重點	五個流程構面	資料輸入、模型運算、結果輸出、人員使用及責任治理。
公共價值路徑必要重點	六項公共價值	正確有效、公平不歧視、透明可解釋、隱私與安全、人類自主、問責與救濟。
PHRM 連結	人事實務情境	可帶入招募甄選、考績、陞遷、訓練或人才配置，維持公共人資定位。

答案段落寫法

每項風險可依「風險如何形成 → 對公務決策或人事決策造成何種影響 → 管理者或機關應注意何種控制要求」展開。

七、解法一：AI 決策流程風險法

(一)解題方向

以「風險發生在哪一個決策環節」為核心問題，依 AI 參與決策的流程逐段盤點，架構直觀且便於完整檢查。

資料輸入	→	**模型運算**	→	**結果輸出**	→	**人員使用**	→	**責任治理**

（二）完整答題架構

大小標題	各段內容與作答重點
一、前言	先交代 AI 可協助公務機關處理大量資料、辨識趨勢及比較政策方案，再指出公務決策涉及人民權益與行政責任，管理者須檢驗 AI 建議並保有最終判斷。兩項內容均應出現，可合併為一段。
（一）AI 協助公務決策之功能	AI 可支援資料整理、預測、方案比較及風險評估，具有提升決策效率與一致性的功能。
（二）公務決策之風險控制要求	公務決策須兼顧正確、公平、透明、程序保障及課責，管理者應查核資料與模型，並由法定職權者作成最終決定。
二、內容	依 AI 參與決策之流程，分別分析資料輸入、模型運算、結果輸出、人員使用及責任治理五個階段的風險。
（一）資料輸入階段	資料品質、代表性、時效及合法利用會決定 AI 推論的基礎，錯誤資料與歷史偏誤可能在後續流程被放大。
1. 資料品質與時效不足	行政資料若錯誤、遺漏、過時或樣本代表性不足，模型將在失真的資訊基礎上推論，造成預測及政策建議偏差，影響資源配置與管理決策的正確性。

大小標題	各段內容與作答重點
2. 歷史偏誤及資料治理風險	既有資料可能反映性別、年齡、族群或地區差異；若用於甄選、考績、陞遷或人才配置，模型可能複製不平等並損害功績與平等機會。大量蒐集或目的外利用個資，也會侵害隱私與資訊自主。
（二）模型運算階段	模型設計、變項選擇及推論機制可能形成黑箱、系統性偏差及情境理解不足。
1. 黑箱與可解釋性不足	複雜模型的權重及推論過程難以理解，管理者不易確認建議形成的原因及錯誤位置，機關也難以履行實質審查與理由說明。
2. 演算法偏誤與情境侷限	代理變項與權重設定可能對特定群體形成不利結果；AI 對法律目的、倫理價值、政策脈絡及例外個案的掌握有限，容易提出不合情境的建議。
（三）結果輸出階段	AI 輸出可能包含錯誤資訊，也可能以標準化分類壓縮個案衡量與救濟空間。
1. 錯誤輸出、生成式 AI 幻覺與不當推論	AI 系統可能產生錯誤預測或不當推論；生成式 AI 另可能生成不存在的事實、錯誤法令或虛構資料。管理者若未查證即採用，錯誤資訊將直接進入決策並擴大行政影響。
2. 過度標準化與說理救濟困難	固定分類可能忽略人民或員工的個別情形，在考績、陞遷或人才配置上造成形式一致而實質不公平；輸出欠缺可理解理由時，當事人也難以陳述意見、爭執錯誤及尋求救濟。
（四）人員使用階段	管理者如何理解與使用 AI，會影響輔助工具是否實質支配決策，以及公務人員的責任與能力。
1. 自動化偏誤與過度依賴	AI 具有快速、量化與一致的外觀，可能使主管在考績、陞遷、人力配置或其他管理決策中，忽略反證、現場資訊與專業經驗，讓 AI 建議取代必要的個案審查。
2. 責任推卸與專業能力弱化	承辦人以「系統判定」作為理由，可能降低事實調查、法令適用及說理程度；長期依賴也會削弱獨立判斷、經驗累積與組織學習。
（五）責任治理階段	機關須處理多方權責、委外資訊不對稱、資安威脅及模型持續監測等治理問題。
1. 權責不明與委外依賴	系統開發、資料提供、操作及核定由不同主體負責時，容易形成責任斷裂；政府欠缺技術能力，也可能無法有效檢驗廠商模型與限制。
2. 資安、監測及退場機制不足	資料外洩、模型竄改及服務中斷會破壞決策資訊；模型效能亦可能隨環境改變而下降，機關須定期檢測、校正並建立停用與退場程序。
三、結論	考場擇一使用。總結式結論重申五階段風險及人工覆核；延伸式結論提出資料治理、模型檢驗、內控及公務人員 AI 能力等制度方向。
（一）總結式結論	概括五階段均可能發生風險，強調 AI 維持輔助定位，由法定決策者審查及負責。
（二）延伸式結論	延伸至機關風險評估、內控、監測、資安、責任追蹤與人員培訓，建立可信賴的人機協作。

八、解法一：完整擬答

下列擬答完整列示兩種結論，考場作答時擇一使用。

參考篇幅：採其中一種結論作答，全文約 1,730 字，約占國考答案卷 2.5 面左右。本文為完整擬答，考場上可依時間與作答重點酌予取捨。

一、前言

人工智慧可協助公務機關處理大量資料、辨識趨勢、預測結果及比較政策方案，提升決策效率與一致性。公務決策涉及公共資源、人民權益及行政責任，管理者採用 AI 建議時，仍須確認資料、模型與輸出的可靠性，維護公平、透明及程序保障，並由法定職權者作成最終判斷。其風險可依 AI 參與決策之流程分析如下：

二、內容

（一）資料輸入階段

1. 資料品質與時效不足

AI 的判斷品質取決於輸入資料。行政資料若有錯誤、遺漏、過時、定義不一致或樣本代表性不足，模型即會在失真的基礎上推論，使預測及政策建議產生偏差，進而造成資源錯置或不當分類。管理者應確認資料來源、更新頻率與代表性，並建立資料清理及品質檢核機制。

2. 歷史偏誤及資料治理風險

既有行政紀錄可能反映性別、年齡、族群、地區或社經條件所形成的歷史差異，模型學習後可能複製並放大不平等；若將此類資料運用於招募甄選、考績評量、陞遷或人才配置，可能使特定群體持續處於不利地位，損害功績原則與平等機會。大量蒐集或整併資料，也可能產生目的外利用、保存過久或外洩等問題。機關應檢驗資料偏誤，並遵守目的特定、最小必要及安全管理要求。

（二）模型運算階段

1. 黑箱與可解釋性不足

部分 AI 模型的變項權重及推論過程高度複雜，管理者難以掌握建議形成理由，也不易確認錯誤來源。公務決策涉及法令適用、行政裁量及人民權益，欠缺可解釋性將影響實質覆核、理由說明及後續救濟。機關應要求符合風險程度的可解釋性，並保存資料、版本及操作紀錄。

2. 演算法偏誤與情境侷限

模型設計、特徵選擇及代理變項可能對特定群體形成系統性不利。公務決策又須考量法律目的、比例原則、政策脈絡與例外情形，AI 對非量化因素及價值衝突的掌握有限。機關應在導入前進行偏誤與效度檢驗，使用中比較不同群體結果，並保留人工調整空間。

（三）結果輸出階段

1. 錯誤輸出、生成式 AI 幻覺與不當推論

AI 系統可能產生錯誤預測、不當分類，或將統計相關誤認為因果關係；生成式 AI 另可能生成不存在的事實、錯誤法令或虛構資料。資料分布改變後，模型也可能持續輸出已經過時的結果。由於 AI 答

案常具有完整與量化的外觀，管理者容易低估其不確定性。重要事實、法令及推論均應由承辦人查核來源、適用情境與可信程度。

2. 過度標準化與說理救濟困難

固定分類與標準化輸出可能忽略人民或員工的個別情形；在人事甄選、考績、陞遷或人才配置中，若僅依模型分數決定，即可能造成形式一致而實質不公平。輸出若欠缺可理解理由，當事人也難以確認資料或推論錯誤，進而影響陳述意見及行政救濟。機關應將 AI 輸出轉化為可理解的行政理由，設置人工覆核、更正及例外處理機制。

（四）人員使用階段

1. 自動化偏誤與過度依賴

AI 具有快速、量化及一致的特性，容易形成科技權威，使主管在考績、陞遷、人力配置或其他管理決策中過度信賴其建議，忽略反證、現場資訊、專業經驗及個案差異，形成自動化偏誤。機關應要求使用者記錄採納或不採納的理由，並在高影響案件中安排第二層人工覆核，維持實質判斷。

2. 責任推卸與專業能力弱化

承辦人或主管以「系統判定」作為理由，可能降低事實調查、法令適用、個案審查及說明程度，並將錯誤責任推向模型或廠商。長期依賴 AI 也會削弱專業判斷與經驗累積。機關應明定法定決策者負最終責任，並透過教育訓練及案例檢討維持人員能力。

（五）責任治理階段

1. 權責不明與委外依賴

AI 系統的資料提供、模型開發、操作、覆核及核定可能由不同主體負責，錯誤發生時容易形成權責分散與歸責不明。政府若欠缺模型檢驗能力，也可能受制於外部廠商。機關應在採購、契約、驗收及營運階段明定權責、稽核權、資料範圍及協力義務，建立可追溯的責任鏈。

2. 資安、監測及退場機制不足

公務資料輸入外部平台或委外系統，可能造成個資、公務機密與政策資訊外洩；資料污染、模型竄改及服務中斷也會破壞決策資訊。模型效能另可能隨法令或社會環境改變而下降。機關應持續監測準確度、偏誤與資安事件，設定警示、停用、備援及退場程序。

三、結論（擇一）

（一）總結式結論

綜上，AI 協助公務決策的風險可能出現在資料輸入、模型運算、結果輸出、人員使用及責任治理各階段。公務機關應維持 AI 的輔助定位，由法定決策者進行實質覆核及最終決定，確保決策品質、人民權益與行政課責。

（二）延伸式結論

面對 AI 逐步進入公務決策，政府應將風險管理納入內部控制與人力資源管理，建立資料治理、模型檢驗、人工覆核、持續監測、資安及責任追蹤機制，並培養公務人員辨識 AI 限制與進行專業判斷的能力，形成可信賴的人機協作。

九、解法二：公共價值治理法

(一)解題方向

以「AI 可能損害哪些公務決策價值」為核心問題，依政府決策應維護的公共價值分類，公共部門特性較為鮮明。

正確有效	→	公平不歧視	→	透明可解釋	→	隱私與安全	→	人類自主	→	問責救濟

（二）完整答題架構

大小標題	各段內容與作答重點
一、前言	先說明 AI 可提升政府資料分析與方案比較能力，再指出公務決策須維護人民權益與民主治理價值，故可依正確、公平、透明、隱私、人類自主及問責等標準評估風險。兩項內容均應處理，可合併為一段。
（一）AI 協助決策的治理意義	AI 可提升資訊處理、預測與決策支援能力，成為政府管理的重要工具。
（二）公共價值作為風險評估標準	政府使用 AI 須符合依法行政、人權保障、程序正當及民主課責，公共價值可作為系統性分析標準。
二、內容	依公務決策應維護的六項公共價值，分析 AI 可能造成的風險。
（一）正確性與行政效能	錯誤、過時或不完整資料，加上模型與情境理解限制，可能產生失真的預測及政策建議，使政府以更快速度作成方向錯誤的決定，增加修正成本並降低行政效能。
（二）公平與不歧視	歷史資料中的差別待遇可能被演算法複製，居住地、學校或職業等代理變項也可能間接對特定群體形成不利結果，影響平等、功績及實質公平。
（三）透明、可解釋與程序正當	黑箱模型使管理者難以理解變項權重與推論理由，也使人民或受評人無法確認不利決定的資料基礎與形成過程。機關須揭示 AI 使用範圍、提出可理解理由，並保障陳述意見與人工覆核。
（四）隱私、資料治理與資安	AI 需要大量資料訓練與運作，可能引發過度蒐集、目的外利用及個資外洩；外部平台、委外系統、資料污染及模型攻擊亦會威脅公務機密與決策資訊完整性。
（五）人類自主與行政裁量	管理者可能因科技權威而過度信賴 AI，忽略反證、個案差異及專業經驗；在人事甄選、考績、陞遷與人才配置中，固定模型也可能壓縮比例衡量、例外處理與實質公平。
（六）責任歸屬與行政救濟	AI 系統涉及資料提供者、開發者、廠商、承辦人及主管，錯誤發生時可能責任不明。機關應明確操作、覆核及核定責任，保存決策軌跡，並提供更正、申訴、人工重新審查及其他救濟管道。
三、結論	考場擇一使用。總結式結論重申六項公共價值風險與人類監督；延伸式結論提出將公共價值轉化為資料治理、透明說明、偏誤檢測、責任及救濟制度。
（一）總結式結論	考場擇一使用。總結式結論重申六項公共價值風險與人類監督；延伸式結論提出將公共價值轉化為資料治理、透明說明、偏誤檢測、責任及救濟制度。

大小標題	各段內容與作答重點
（二）延伸式結論	延伸至偏誤檢測、資料治理、透明揭露、人工覆核、資安、責任歸屬與救濟，使科技應用符合依法行政與民主課責。

十、解法二：完整擬答

下列擬答完整列示兩種結論，考場作答時擇一使用。

參考篇幅：採其中一種結論作答，全文約 1,425 字，約占國考答案卷 2 面左右。本文為完整擬答，考場上可依時間與作答重點酌予取捨。

一、前言

人工智慧可協助公務機關整理大量資訊、預測趨勢及比較政策方案，提升管理決策的速度與分析能力。政府決策涉及公權力、公共資源及人民權益，除行政效能外，尚須維護公平、透明、隱私、人類自主與行政課責。因此，可依公務決策應遵循的公共價值，分析 AI 協助決策的風險如下：

二、內容

（一）正確性與行政效能風險

公務決策首先須建立在正確資訊及合理推論上。AI 輸出高度依賴資料品質與模型設定，資料錯誤、遺漏、過時、樣本不足或模型設計不當，均可能造成不可靠的預測與建議；AI 對法律目的、政策脈絡及特殊個案的理解也有限。管理者若直接採用輸出，可能以更快速度作成方向失準的決定，增加資源錯置與後續修正成本。機關應進行資料驗證、效度測試及人工覆核，持續檢查模型在實際情境中的表現。

（二）公平與不歧視風險

公務決策須兼顧平等對待與實質公平。歷史資料可能反映既有社會差異及行政偏誤，演算法學習後會延續甚至放大不平等；居住地、學校、職業或消費紀錄等表面中立的代理變項，也可能間接反映性別、族群或社經身分。機關若將模型運用於甄選、績效評估或資源分配，可能使特定群體持續處於不利地位。管理者應檢驗不同群體的輸出差異，確認變項與決策目的具有實質關聯，並保留個案調整機制。

（三）透明、可解釋與程序正當風險

公務決策須提出理由，並保障受影響者理解、陳述意見及接受人工審查的機會。複雜模型的變項權重及推論過程可能難以理解，機關因而無法充分檢驗 AI 如何形成建議，也難以確認錯誤來源。人民、應考人或受評人若不知道 AI 的使用情形、資料基礎及重要判斷因素，便難以指出錯誤或表達個案情形，程序公平與政府信任亦會受到影響。機關應揭示 AI 使用範圍，保存可追溯紀錄，以一般人可理解的方式說明理由，並提供陳述意見及人工覆核程序。

（四）隱私、資料治理與資安風險

政府掌握大量個人及業務資料，使用 AI 時須維護隱私、資訊自主與資料安全。大量蒐集可能超出特定目的所需，資料整併、二次利用及長期保存亦可能增加當事人無法掌握資料流向的風險。機關使用外部平台或委外系統時，個資、公務機密與政策資訊另可能外洩；資料污染、模型竄改及服務中斷也會影響決策資訊的完整性與可用性。故應採取最小必要蒐集、分級授權、去識別化、資安檢測及事件通報措施。

（五）人類自主與行政裁量風險

公務決策中的人類自主，包含管理者保有專業判斷，以及人民或員工不受純粹自動化結果支配。AI 輸出具有量化與科技權威的形象，容易使管理者產生自動化偏誤，降低對反證、現場資訊、個案差異及專業經驗的重視。在人事甄選、考績、陞遷或人才配置中，固定模型也可能將複雜情境化約為單一分數，壓縮比例衡量、例外處理與實質公平的空間。機關應明定 AI 僅具輔助性，要求法定決策者進行實質審查，並保留人工介入、調整及停止使用的權限。

（六）責任歸屬與行政救濟風險

AI 系統涉及資料提供者、開發者、委外廠商、承辦人、主管及機關首長，錯誤發生時容易形成責任分散與歸責不明。技術服務由廠商提供，也不會取代機關及法定決策者的行政責任。機關應明確各階段操作、覆核及核定責任，保存資料來源、模型版本與決策軌跡，並設置資料更正、申訴、人工重新審查及其他救濟管道，使受不利決定者得以有效主張權利。

三、結論（擇一）

（一）總結式結論

綜上，AI 協助公務決策可能同時影響正確有效、公平不歧視、透明可解釋、隱私與安全、人類自主及問責救濟等公共價值。政府應保留人類監督與最終判斷，使科技運用符合依法行政、權利保障及民主課責。

（二）延伸式結論

AI 治理應將公共價值轉化為具體制度。政府宜建立偏誤檢測、資料治理、透明說明、人工覆核、資安管理、責任歸屬與人民救濟機制，並透過教育訓練提升公務人員 AI 素養，使科技創新與公務決策正當性相互結合。

十一、自行練習後的作答檢查

自行完成答案後，可依下列項目自評。本段供練習後檢查，不必全部寫入考場答案。

1. 前言是否點出 AI 的輔助功能與公務決策的風險控制要求？

2. 內容是否形成一套清楚且前後一致的分類架構？

3. 是否處理資料或模型限制、錯誤輸出、過度依賴、公平透明隱私、人工覆核與行政責任？

4. 是否連結至少一項 PHRM 人事決策情境，例如甄選、考績、陞遷或人才配置？

5. 每項風險是否說明形成原因及對決策品質、人民權益或人事公平的影響？

6. 結論是否回到 AI 輔助定位、法定職權者最終判斷及機關內控？