人工智慧管理系統

ISO 42001
概覽指南

內容

遨遊人工智能領域

瞭解重塑技術、風險和責任的趨勢

人工智慧 (AI) 越來越多地被應用在所有運用資訊科技的領域，並預計將成為主要的經濟動力之一。這一趨勢的結果是，某些應用在未來幾年可能會引起社會挑戰。從金融和製造業到醫療保健和物流，AI 正在推動自動化、效率和決策方面的進步。生成式人工智能和機器學習的廣泛應用釋放了新的可能性 – 但也加速了對更強大的 **治理**、更高的 **透明度** 和更清晰的 **問責性的 需求**。

全球 AI 領域正受到三種趨勢的影響：

加速的實施

企業正在快速地將人工智能嵌入到核心業務中。根據 IBM 的資料，42% 的企業已經在探索或積極部署生成式 AI。然而，這種速度往往趕不上正式治理結構的發展 – 在監督、品質保證和風險管理方面造成缺口。

不斷演進的法規

各國政府和監管機構正在通過新的框架來應對日益增長的社會關切，以確保人工智能系統的安全、公平和可解釋性。2024 年通過的歐盟 AI 法案開創了以風險為基礎的監管先例，類似的舉措也正在全球範圍內進行。這個訊息很清楚：對於 AI 的信任必須是贏得的，而不是假設的。

不斷提高的審查和道德期望

從資料隱私、智慧財產權到偏見與責任，各機構都面臨壓力，必須證明其使用的人工智慧是合乎道德、負責任且安全的。公眾的信任是脆弱的 – 而失策可能會迅速導致聲譽受損、法規後果和機會流失。

這些趨勢標誌著一個轉捩點。AI 已從試點計畫轉型為策略性基礎架構。 而隨著這一轉變，我們需要結構化、全系統的治理，能夠隨著雄心壯志而擴展，並經得起審查。

什麼是 ISO 42001?

第一個國際人工智慧管理系統標準

ISO 42001 是全球第一個專門針對人工智慧的管理系統標準 – 提供一個結構化的架構，協助組織以負責任的態度管理人工智慧。

本標準專為開發、部署或依賴人工智慧的組織所設計，列出建立、實施、維護及持續改善人工智慧管理系統 (AIMS) 的要求。

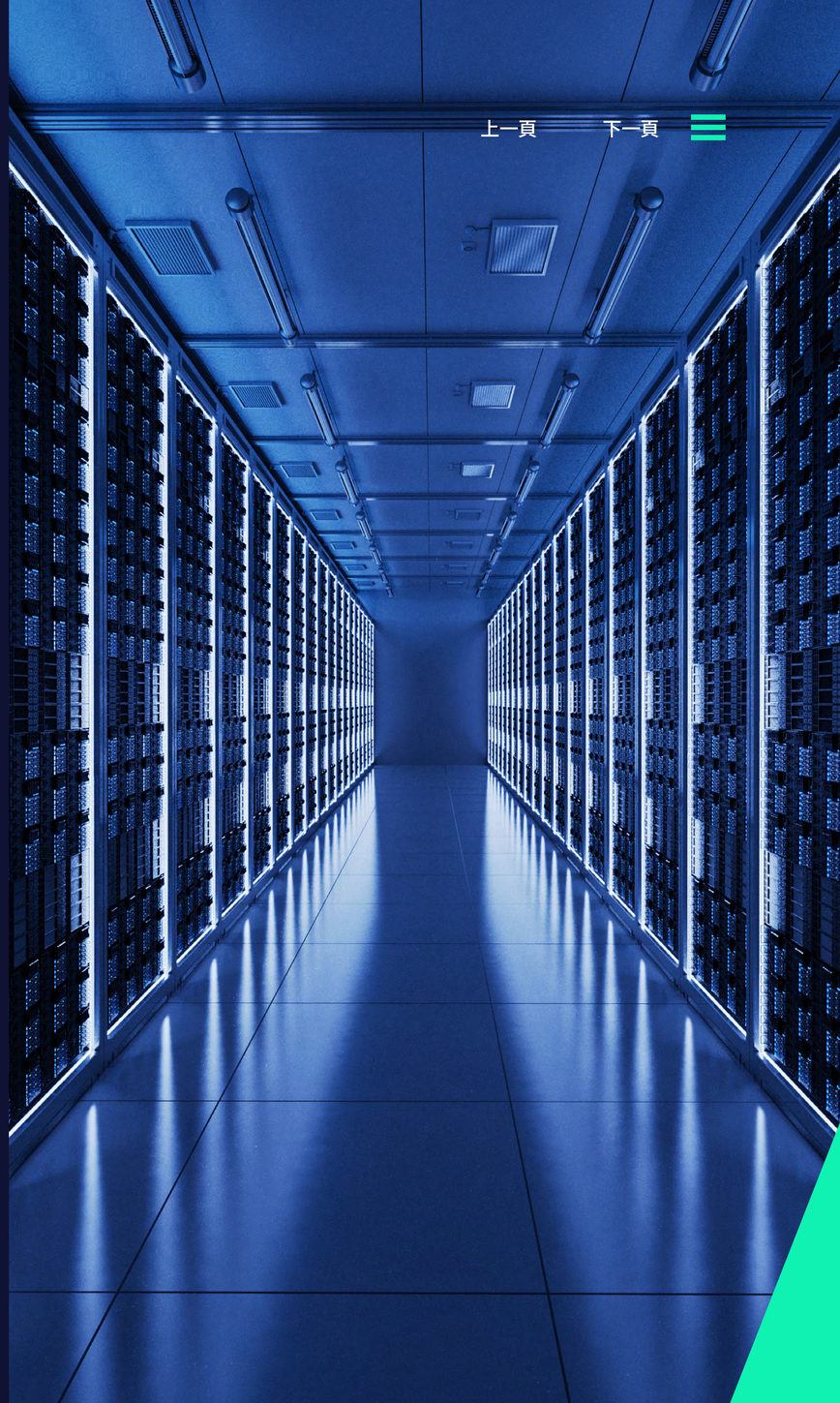
無論您是在業務運作中整合人工智慧，或是提供人工智慧產品和服務，ISO 42001 都能協助您納入道德原則、管理風險，並符合全球對值得信賴的人工智慧的期望。

標準的主要內容包括

- **治理與問責** – 定義 AI 相關活動的角色、責任與監督
- **基於風險的控制** – 在整個 AI 系統生命週期中識別和處理風險
- **透明度與可解釋性** – 支援清楚溝通 AI 系統如何運作及做出決策
- **持續改善** – 運用回饋、監控和績效資料，不斷強化您的 AIMS

ISO 42001 遵循 Annex SL 結構，可輕鬆與其他標準 (例如 ISO 9001、ISO 27001 或 ISO 45001) 整合，以一致且協調的方式管理整個組織的風險。

透過採用 ISO 42001，組織可以展現出對負責任創新的明確承諾，在日益以 AI 為主導的世界中，與客戶、合作夥伴和監管機構建立信任。



為什麼要取得驗證?

ISO 42001 驗證不只是一種形式。
它是一個獨立、全球認可的標誌
可保證您的組織正以負責任的態度
管理人工智慧 並且行之有效。

顯示對最佳實務的承諾
驗證標誌著您對於建立道德、透明且管理完善的 AI 系統的認真態度 – 符合國際期望。

與客戶建立信任和信譽
對許多產業而言，認證正逐漸成為一種交易許可證。它可以讓客戶、合作夥伴和利害關係人放心，您的人工智慧實務是安全、負責任且管理完善的。

支持隱私、道德與資訊安全目標
驗證程序有助於將 AI 治理嵌入到您更廣泛的風險管理系統中 – 加強資料保護、負責任的創新和道德監督。

為您的合規性提供未來保障
隨著全球 AI 法規的加速發展，ISO 42001 驗證可協助您的組織領先不斷演進的法律和產業要求 – 降低風險並支援長期的適應力。

ISO 42001 主要要求：

附件 SL 條款結構

ISO Annex SL 架構包含十個條款。管理系統標準（包括 ISO 42001）中的所有內容都必須符合所有十個條款的標準，才能遵循 Annex SL 架構。這些條款可分為以下幾類：

第 1 條
範圍

定義 AI 管理系統的預期成果及其在組織內的適用性。

第 2 條
規範參考

列出對應用 ISO 42001 必不可少的參考標準。

第 3 條
術語和定義

提供整個標準中使用的核心術語定義，以確保共同理解。

第 4 條
組織背景

考慮內部和外部因素、利害關係人的期望，以及 AI 系統的使用範圍。

第 5 條
領導能力

概述了最高管理層在政策資源調配和推廣負責任的 AI 文化方面的責任。

第 6 條
規劃

處理組織如何識別與回應 AI 風險與機會，以及設定 AI 目標。

第 7 條
支援

涵蓋有效管治所需的資源、能力、意識、溝通和文件。

第 8 條
操作

定義如何實施 AI 相關控制，包括風險評估和系統影響評估。

第 9 條
績效評估

需要監控、測量、內部稽核和管理審查來評估系統效能。

第 10 條
改進

詳細說明組織持續改善、不符合處理和糾正行動的方法。

在大多數情況下，這些條款使用相同的核心文字，不論它們適用於哪個標準，並分享共同的術語和定義，以促進各管理系統標準的一致性和相容性。對於 ISO 42001 而言，這可確保 AI 治理嵌入熟悉且行之有效的管理系統架構中。

在人工智能的背景下應對 ISO 42001 的要求

與其他 ISO 管理系統標準一樣，ISO 42001 也是圍繞第 4 條至第 10 條而建立，涵蓋了情境、領導、規劃支援和持續改善等領域。ISO 42001 的獨特之處在於如何針對與人工智慧相關的挑戰和風險，量身打造這些耳熟能詳的概念。

以下各節總結了這些條款在人工智慧環境中的應用 – 從道德治理和資料透明度到風險管理和生命週期監督。

組織背景

ISO 42001 要求組織透過了解其內部與外部環境來界定其人工智慧管理系統的範圍。這包括法律義務、特定產業的 AI 風險、文化與道德規範、利害關係人的期望，以及組織在 AI 生命週期中的角色 – 無論是身為開發者、部署者或使用者。這些洞察力會影響 AI 的管理方式，並有助於決定哪些控制是相關的。此外，我們也鼓勵組織評估更廣泛的社會議題 (例如氣候變遷或數位公平) 應否影響其方法。

領導與治理

ISO 42001 的核心要求之一，是領導階層對負責任的人工智慧使用做出明顯且持續的承諾。高級管理階層必須建立明確的人工智慧政策、設定與組織價值相符的目標，並確保在整個人工智慧生命週期中定義角色與責任。這包括持續監督風險評估、影響評估和 AI 的使用，尤其是在敏感或高風險的情況下。最高層的積極參與有助於確保治理貫穿整個企業，而非事後的想法。

道德與透明度

道德考量貫穿始終
ISO 42001。組織必須證明他們如何評估和處理人工智能對個人和社會的潛在影響。這包括考慮公平性、非歧視性和人類自主性，以及資料隱私和結果透明度。控制措施必須到位，以識別和緩解意外後果 – 這些做法的證據必須記錄、監控和定期更新。

風險與機會管理

組織必須評估和處理特定的 AI 風險，包括那些影響合規性、聲譽和安全的風險，同時也要找出創新和改善的機會。ISO 42001 承認各行各業的風險承受能力和定義可能有所不同，因此該架構具有適應性。重要的是，必須有意識地在政策的指導下做出決策，並隨著時間的推移評估其有效性。

持續改善

ISO 42001 遵循「計畫-執行-檢查-行動」(Plan-Do-Check-Act，PDCA) 模式。這表示持續改善並非可有可無，而是根深蒂固的。組織必須監控其 AI 系統效能、進行內部稽核，並定期檢閱治理流程。無論是矯正行動或適應法規變革，目標都是持續改善 AI 管理系統的適用性、充分性和有效性。

ISO 42001 附件 A 控制

將負責任的 AI 轉化為行動

除了 10 個主要條款之外，ISO 42001 還在附件 A 中包含一組支援控制目標。這些控制目標旨在協助組織實施實際、可稽核的措施，以支援可信賴的人工智慧開發與使用。

附件 A 控制項涵蓋 10 個類別：

- **與 AI 相關的政策** – 確保明確的領導意圖與方向
- **內部組織** – 定義管理角色和責任
- **資源** – 管理用於 AI 的工具、資料和基礎設施
- **影響評估** – 評估對個人和社會的風險
- **生命週期監督** – 使系統設計符合道德與法規目標
- **AI 系統的資料** – 確保品質、來源和相關性
- **為相關各方提供資訊** – 提升透明度與問責性
- **AI 系統的使用** – 管理範圍、意圖和保障措施
- **第三方關係** – 設定對供應商和合作夥伴的期望
- **文件和可追蹤性** – 支援可解釋性及可稽核性

這些控制領域提供了實施 ISO 42001 的實用基礎。它們可適用於不同層級的 AI 成熟度和複雜性 – 並成為就緒評估和驗證的關鍵部分。

實施 ISO 42001

實作 AI 管理系統的重點領域

ISO 42001 提供了一個使用「計畫-執行-檢查-行動」(Plan-Do-Check-Act，PDCA) 模式來實施負責任和有效的 AI 管理實務的框架 – 這是所有以附件 SL 為基礎的 ISO 標準所共有的核心基礎。

本標準透過相互連結的管理流程，支援人工智慧管理系統 (AIMS) 的開發，包括

- **意識**

  組織應該訓練團隊，讓他們了解人工智慧的好處、風險與道德考量，從開發人員到決策者，都應該建立這樣的意識。

- **責任**

  必須指派明確的角色與職責，確保 AI 治理被理解並嵌入整個系統生命週期中

- **回應**

  應該及時採取協調的行動來管理風險、應對 AI 系統的影響，並在必要時採取矯正措施。

- **風險評估**

  與 AI 系統相關的風險–包括技術故障、非預期結果或濫用 – 必須以結構化、循證的方式進行評估。

- **系統設計與開發**

  AI 系統的開發與部署必須符合文件化的政策與流程 – 支援道德使用、透明度與生命週期責任。

- **治理與控制**

  各機構應採用全面的方法來管理人工智能，涵蓋資料品質、可解釋性、公平性、可靠性和人為監督。

- **重新評估與持續改善**

  作為 PDCA 模型的一部分，應定期監控和審查績效。內部稽核和管理審查有助於確保 AIMS 繼續達成目標，並適應新出現的風險和法規。

使用 ISO 42001 建立您的路線圖

為有效的 AI 治理奠定基礎

實施 ISO 42001 始於明確的目標和強大的領導支持。這些早期步驟有助於確保您的人工智慧管理系統 (AIMS) 既實用又符合組織的目標。

1. 獲得領導階層的承諾

資深管理階層必須領導 AIMS – 設定明確的目標，界定組織的成功樣貌。無論是加強監督、滿足法規期望或建立信任，領導階層都應該分配所需的資源和方向，將負責任的 AI 使用植入整個企業。

2. 瞭解您的 AI 環境

評估 AI 在組織內的開發、部署或使用方式。考慮法律義務、特定產業的風險和利害關係人的期望。此瞭解應能為您的資源配置提供參考 – 包括負責任、有效管理 AI 所需的時間、技能和預算。

3. 評估訓練需求

AI 治理涵蓋技術、道德、法律和作業領域。讓跨功能團隊參與，並確保提供量身打造的訓練 – 從更廣泛團隊的認知到專家的稽核訓練。建立跨角色的能力是有效、可擴展實施的關鍵。

4. 定義 AIMS 的範圍

澄清您的 AIMS 將涵蓋哪些團隊、系統和地理位置 – 包括內部開發的 AI、第三方工具和高風險使用個案。明確定義的範圍可讓您的管理更專注且符合目的。

5. 規劃並執行您的 AIMS

制定實際的實施計畫，涵蓋時間表、責任和所需資源。引入必要的控制、文件和管理流程。建立定期審查、回饋環路和績效監控，以確保您的 AIMS 與您的業務和 AI 環境同步發展。

6. 進行差距分析

檢閱您目前的管治、風險與合規架構，以找出您的組織在哪些方面已經符合標準，以及在哪些方面存在缺口或弱點。這些洞察力將有助於優先改善和避免重複。

7. 預約您的認證稽核

一旦您的 AIMS 就位，請安排 LRQA 進行 ISO 42001 驗證稽核。我們的兩階段流程將評估您的系統設計和實施 – 協助您向利益相關者展示責任感、誠信和負責任的創新。

8. 嵌入持續改善和回應

成熟的 AIMS 不僅包括初始實施 – 它需要持續學習和改進的機制。建立回饋機制、定期稽核和績效指標，以適應不斷演進的 AI 風險和技術。實施特定於 AI 的事件回應計畫，以確保快速、透明且有效地處理錯誤、偏差或系統故障。

將 ISO 42001 與現有的管理系統整合

以現有成果為基礎。加強您對 AI 和資訊安全的管理。

對許多組織而言，ISO 42001 並非起點 – 而是自然延伸。如果您已經擁有 ISO 27001 認證，您就有能力有效率地整合 ISO 42001。

ISO 27001 提供了一個行之有效的資訊安全風險管理架構，其關於資料機密性、完整性和可用性的控制措施直接支援 ISO 42001 的許多要求。透過結合這兩項標準，您可以建立統一的治理方法，同時處理資訊與 AI 相關的風險。

為什麼整合是合理的

共享結構

ISO 42001 遵循 Annex SL 架構 – 與 ISO 27001, ISO 9001、ISO 45001 及其他組織所使用的高階架構相同。這可讓政策、領導、規劃、運作和評估保持一致。

有效利用資源

整合有助於減少稽核、文件和內部審查的重複。團隊可以統一報告、目標和控制 – 簡化合規性並改善監督。

一致的風險管理

這兩項標準都要求採用基於風險的方法 ISO 27001 專注於資訊資產，而 ISO 42001 則延伸至人工智慧系統，包括如何使用資料來訓練、驗證和 操作這些系統。

更強的系統彈性

整合式方法更容易發現系統問題、全面解決問題，並向客戶、監管人員和合作夥伴展示聯合治理模式。

我們的 ISO 42001 服務

訓練

透過一系列針對不同經驗等級所設計的課程，建立您對 ISO 42001 的知識 – 透過多種學習方式來傳授。

差距分析

這是一項可選服務，我們的專家稽核人員會在認證前幫助您找出系統中的任何關鍵、高風險或薄弱區域。

驗證

我們根據 ISO 42001 的要求評估您的 AI 管理系統 (AIMS)。

綜合評估

如果您已實施多重管理系統，您可以從整合式稽核與監督方案中獲益，這是一種更有效率且更具成本效益的風險管理方式。

檢視可用的課程

為什麼與我們合作?

在 LRQA，我們協助組織開發穩健、適應未來的風險管理方案，以安全、負責任且有效地採用人工智慧與技術。

從確保符合不斷演進的人工智慧法規，到強化資料安全性以及在治理中嵌入最佳實務，我們都能提供滿懷信心推動創新所需的保證 – 協助企業整合人工智慧與技術，同時主動管理風險。

現場專業知識

我們的解決方案由專精於網路安全、合規性和供應鏈風險管理的全球專家團隊提供，協助您掌控 AI 相關風險、滿足不斷演進的法規要求，並將負責任的 AI 實務整合至您的營運中。

以解決方案為基礎的夥伴關係

我們不僅提供驗證，還與您並肩合作，將 AI 治理融入您更廣泛的風險與合規策略中。我們量身打造的方法可確保人工智慧與技術在滿足不斷演進的法律與道德期望的同時，協助推動企業永續成長。

持續保證

AI 驅動的風險需要持續監督。我們的即時風險管理方法可主動解決問題，減少業務中斷並加強應變能力。我們連接的風險管理解決方案組合可協助企業超越法規要求，並將 AI 風險管理嵌入日常營運中。

資料驅動的決策

我們利用數位平台和分析，深入洞察您企業的 AI 風險。我們的人類智慧透過資料驅動的工具加以強化，可協助組織找出弱點、預測未來風險，並有信心地做出明智決策。

關於 LRQA:

LRQA在評審、諮詢、檢驗和網路安全服務方面積累了數十年極為豐富的專業知識。我們基於解決方案的合作夥伴關係以資料 驅動的洞察力為後盾，幫助我們的客戶解決他們最大的業務挑戰。

LRQA在全球150多個國家/地區運營，擁有5,000多名員工；我們的合規、供應鏈、網路安全和ESG專家屢獲行業殊榮幫 助幾乎遍佈所有行業的61,000多家客戶預測、減輕和管理風險，無論其在哪裡運營。

我們很自豪成為世界領先的保障服務合作夥伴之一。在所有領域，我們都致力於為我們的員工、客戶、社區和地球塑造 更美好的未來。與我們合作，讓我們一起共創美好未來。

聯繫我們

網站: www.lrqa.com/zh-tw
電話: (02) 27166085

扫码关注官方微信