人工智慧與公正文化原則

蕭光霈 譯

歐洲執委會(European Commission)針對人工智慧提案建立一項法律規範框架。鑒於某些風險與契機，本文作者強調以「公正文化」(Just Culture)調和人工智慧運用之重要性，確保決策、標準、訓練與責任歸屬能夠清楚明瞭。

本文重點：

- 歐洲執委會提案建立運用人工智慧的統一規定，以突顯其潛在利益與競爭優勢。
- 提案強調在設計、發展高風險的人工智慧系統，尤其是運用在與安全息息相關之環境中，其作業透明、韌性與接受人類監督之必要性。
- 將人工智慧運用在航空業，會產生責任歸屬與下達決策之問題，我們需要對於人類與機器間責任分擔之觀念有所轉變，避免單方面對人類操作者施加過多負荷。
- 人工智慧的引進，對於我們過去檢驗意圖與因果關係之作法形成考驗，建議對責任歸屬建立一套滑尺制度(sliding scale system)，以因應人工智慧的獨特性質，採取持平之作法。

■ **為達成「公正文化」原則，我們必須就人、機之間相對關係，思考人類行為、人員訓練與標準規範，確保在人類監督與人工智慧能力發揮之間採取均衡作法。**

歐洲執委會於 2021 年 4 月提案針對人工智慧制定統一規定。這項尚未在歐洲議會 (European Parliament) 進行表決之草案，目的在於強調這項技術之重要性，並據草案指出，人工智慧可以「有助於對社會、環境構成有利成果，並對公司企業與歐洲經濟提供關鍵的競爭優勢」。

人工智慧將能藉由改進預測、優化作業與資源分配，以及提供個人化的服務，達成前述之目標。

依據提案，人工智慧的定義是針對人類定義目標的特定組合，產生輸出之軟體。這些輸出包括任何形式的內容 (contents)、預測與建議，或是能對與其互動之環境構成影響的決定。

◎ *基於風險建立之作法*

提案建立規範人工智慧之規則，係採基於風險建立之作法所擬，側重於在產品上構成安全組件之系統。目標在於將這些規則與現今相關行業安全法規整合，以取得一致。

在航空業中，當人工智慧系統運用在或就是在某產品中，成為發揮安全功能之「安全組件」一部分時，在某些程度上，視為是間接受到此項歐盟提案影響之高風險環境。人工智慧系統發生故障或功能異常，將會對個人與財產構成危害。

有鑒於此，在航空業領域中引進人工智慧，應要遵守此項提案所提之某些原則。其中某些條款對於安全極為重要。

首先，是項提案指出，設計與研發高風險的人工智慧系統時，應要確保其作業透明，足以讓使用者理解其系統輸出結果，並適當運用。

提案寫道，針對系統內部或是其操作環境，尤其是因為其與人類或其他系統互動時，可能發生之錯誤、故障或矛盾，高風險的人工智慧系統應具備一定之韌性。

另外，提案中表示，應透過人、機介面工具的視角來設計與研發人工智慧，且人工智慧運用期間，應接受所謂「自然人」(natural persons，譯註：具有自然生命的個人) 的監督。在此項條款中，人類監督工作之特定目標，係當高風險人工智慧系統依其原訂用途使用時，或是預判所及之合理誤用情況下，避免或儘可能減少恐會對個人健康、安全或基本權利構成風險。

◎ *人類角色*

提案亦述及，人類監督工作，必須完全瞭解人工智慧系統之能力與限制，並且要能適時監視其作業情況，以發現並解決任何異常與失能之徵狀。

為此項監管草案調整的說法是，監督作法應「讓奉派擔任監督的個人，在適當情況下，應遂行以下事項：

(一) 注意並適切瞭解高風險人工智慧系統的相關能力與限制，並能適時監視其作業，因此異常、失能與不預期的操

作舉動才能及早發現與解決。

(二) 持續注意對於高風險人工智慧系統之輸出，自動產生依賴性或過度依賴的可能傾向(「自動化偏差」，automation bias)，尤其是針對用於提供決策建議供自然人運用之高風險人工智慧系統。

(三) 要能正確解讀高風險人工智慧系統的輸出，尤其要考量該系統之特性以及可用的解讀工具與作法。

(四) 要在任何特定情況下，能夠決定不使用人工智慧系統，抑或無視、否決或逆轉高風險人工智慧系統的輸出結果。

(五) 除非是在人為干預會升高風險，或是人為干預會對公認最新科技構成負面影響的情況之外，要能夠介入高風險人工智慧系統作業，或是能透過按下「停止」鈕或某個類似程序干預系統，讓其在安全狀態下暫停運作。

(歐洲議會於 2023 年 6 月 14 日核定以上第 (一) 至 (五) 項，而此五項與原本措辭用語有所不同。)

◎ 人類監督與責任歸屬

以航空業人工智慧為主題的研究文件與報告，皆有一共同項目：「以人為本之作法」(human-centered approach)。這些文件中包括國際民航組織 (ICAO)2019 年論及航空業的人工智慧與數位化之工作報告、歐洲航空 / 飛航管理 (European Aviation/ATM) 人工智慧高層工作群 (AI High Level Group) 的《以人工智慧飛行》(Fly AI) 報告 (EUROCONTROL, 2020)、歐洲航空安全局 (EASA) 的《人工智慧發展路線圖》(Artificial Intelligence Roadmap, 2020)，以及歐洲單一天空飛航管理研究 (SESAR) 中之《歐洲飛航管理主計畫》(European ATM Masterplan, SESAR Joint Undertaking, 2020)。相關規定在規劃時，皆瞭解所有作業與活動均由人類執行。

但是，這項有關建立人工智慧法規的提案，似乎是從以人為本的作法，轉變為以人類進行監督之作法。因此產生不同的問題。

在航空業環境中引進人工智慧，可能涉及包括人員、航空公司、空中導航服務供應商 (air navigation service provider, ANSP)、國家，以及製造商等數種角色。如《國際民航公約第 11 號附約》(ICAO Annex 11)(亦如第 9426 號與第 4444 號文件) 與歐盟之歐洲單一天空法規彙編 (2018 年第 1139 號條例) 及飛航管理認證與人員授證規定等現行法規，業已考量飛航管制員之觀點。

從責任歸屬的角度，在航空業 (如同其他產業) 使用人工智慧，涉及多種責任類型，包括刑事、民事 (合約內與合約外之責任)、國家 / 行政、產品、組織與代理等類型之責任。

◎「黑箱問題」

提案中針對人工智慧的定義與框架協議，以及人類擔負之責任 (就監督與「維護之責」而言)，其理解的角度，應從人工智慧係透過神經網路，將問題細微分解，然後循線性方式逐步解決，以發揮其功能。但我

人工智慧的功能展現，對過去幾近全般法律領域，用以檢驗意圖與因果關係之作法形成考驗

們無法確實瞭解其演算法處理為何，或其所採方法為何。此即稱之為「黑箱問題」，因為人工智慧可視為是一個黑箱，無法窺探其內部運作。

「相關規定在規劃時，皆瞭解所有作業與活動均由人類執行。但是，這項有關建立人工智慧法規的提案，似乎是從以人為本的作法，轉變為以人類進行監督之作法。

人類可以輸入內容與目標，讓人工智慧運作 (以「黑箱」方式進行)，但需監督其發揮功能，必要時插手干預。但是，回想起來還是有倫理上問題：人類決定插手干預之依據為何？人工智慧能夠建立衡量人類行為的標準或基準嗎？因此可產生兩種情境：

一、 人工智慧建議採取一項正確行動，但飛航管制員並未採納而導致發生事端：

- 飛航管制員對於應盡而未盡專業職責之疏失，是否應該負責？
- 人工智慧建議採取「正確行動」之依據為何？人工智慧是否採取與飛航管制員不同的標準或基準？
- 飛航管制員有責任要採納人工智慧的建議嗎？
- 人工智慧的建議可以做為證據嗎？

二、 人工智慧建議採取一項錯誤行動，而飛航管制員採納其建議而導致事端：

- 飛航管制員對於應盡而未盡專業職責之疏失，是否應該負責？
- 飛航管制員對於人工智慧發揮功能的方式，是否具備恰當之心理模式？

◎ 人、機互動

為配合此一框架並解決前述之問題，同時又能遵守公正文化之原則，從人、機關係之角度檢視人類行為與訓練相當重要。我們必須澄清何者才能下達決策、下達決策的時刻與原因，以及其所依據之標準與訓練為何。

對處於人類與機器互動頻繁交錯模式運作的情況下，此點格外重要。其目的應在於減低對於機器的過度信賴，以及其他意外結果。

引進自動化，讓工作與責任逐漸交由科技接管，造成損害之責任理當從人類操作員手中，轉由負責設計、研發、運用、整合與維護這項科技產品的組織來承擔。然而，人工智慧之功能展現，對過去幾近全般法律領域，用以檢驗意圖與因果關係之作法形成考驗。這類型評估預測結果與決策基礎之檢驗作法，運用在黑箱作業的人工智慧，恐無法發揮效用。

解決此問題之作法，不應針對人工智慧以某種特定透明的標準，實施嚴格之責任歸屬與監管框架。反之，採取調適目前對於意圖與因果關係檢驗之彈性制度，可以產生某種更適切作法。因此，當人工智慧自主運作或欠缺透明度的情況下，此舉會影響對於當下責任歸屬之要求。從另一方面來說，當由人類監督人工智慧或人工智慧運作一切透明的情況下，此舉維持過去的意圖與因果關係檢驗作法。

◎ 公正文化與人工智慧

迄今，我們對於機器運作採取之作法，係由一單純原則所引導：我們知道輸入為何，瞭解其運作方式，也知道預期的輸出的產物。此舉讓我們目光放在人類對於錯誤、疏失與偏差之考量。

引進人工智慧後，我們恐必須處置機器出錯的問題。若將所有責任全都歸咎於人類與其監督職責，這是不公平、錯誤，甚至悖於工作倫理。

觀念轉變非但在事後檢討，律定責任歸屬或進行安全評估時相當重要；且在事前必須實施預防與預警作為時，亦同樣重要。此作法有助於深化「公正文化」原則，此原則不應修改，但須考量將人工智慧視為整體過程中之參與者。

參考資料：

EASA (2023, May 10). EASA artificial intelligence roadmap 2.0: A human-centric approach to AI in aviation. https://www.easa.europa.eu/en/document-library/generalpublications/easa-arti cial-intelligence-roadmap-20

EUROCONTROL (2020, March 5). The FLY AI report: Demystifying and accelerating AI in aviation/ATM. https://www.eurocontrol.int/sites/default/files/2020-03/eurocontrol-fly-ai-report-032020.pdf

European Commission (2021, April 21). Proposal for a regulation laying down harmonised rules on artificial intelligence. https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-armonisedrules artificial-intelligence

ICAO (2019, January 8). Working paper: Artificial intelligence and digitalisation in aviation, 2019. https://www.icao.int/Meetings/a40/Documents/WP/wp_268_en.pdf

SESAR Joint Undertaking (2020, December 17). European ATM master plan 2020.

作者簡介：

Federico Franchina 現職為墨西拿大學(Universit degli Studi di Messina)海空運輸法教授，曾任歐洲空中飛航安全組織(EUROCONTROL)法律專家與義大利建設暨運輸部公共工程高階會議之專家。

譯自 Hindsight 35 | Summer 2023