

人工智慧與聘僱歧視問題簡析 ——以性別歧視為核心

洪瑩容 | 中興大學法律專業學院 助理教授



◎ 壹、前言

隨著數位時代之來臨，越來越多企業使用人工智慧進行聘僱程序，其優點在於能迅速有效率處理大量申請文件，大幅降低時間、金錢成本之耗費。此外，相較於傳統冗長、不透明之招募程序，由人工智慧與演算法來篩選合適之求職者，常被認為較能對於求職者進行客觀評價，蓋其可屏除人類主管因不同的生活經驗、成長背景而產生對於「性別」之偏見或刻板印象，

理論上應有助於消除「就業性別歧視」。然而，自有人工智慧技術以來，關於人工智慧歧視之爭議案例屢見不鮮¹。隨著企業將越來越多的人力資源工作交由人工智慧系統處理，也必須面對使用此類技術可能會導致就業性別歧視的問題，此是否會因雇主使用人工智慧系統在判斷上而有所差異，有待進一步之討論²。

1. 舉例言之，在英國的演算法中，若求職人的姓名為「Jared」並且在高中打過曲棍球，演算法就會建議公司僱用。Vgl. Waas, KI und Arbeitsrecht, RdA 2022, 125, 126.
2. 礙於篇幅限制，在此僅簡要說明適用現行法律下可能產生之問題，本文之論述主要節錄自洪瑩容，企業使用人工智慧之聘僱歧視問題研究——兼論對於現行反歧視法制之檢討，興大法學，第34期，2023年11月，頁1-75一文，相關問題之詳細之分析與論述，可參見此文。

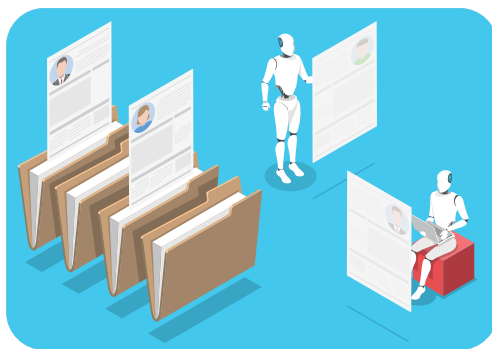
◎ 貳、人工智慧之聘僱歧視

針對企業在聘僱程序中導入人工智慧之形式以及可能產生之歧視問題，茲分述如下：

一、為了提高人工智慧系統辨識優秀人才之能力，企業可以採取的作法為：將過去優秀員工之資料作為人工智慧系統之訓練數據，供人工智慧學習建立適用於該企業之模型³，然而，當企業所提供之資料不夠全面，即可能使演算法產生偏誤而導致歧視之結果⁴。舉例言之，企業過去所僱用之男性員工多於女性，且男性員工擔任主管職缺之比例高於女性，當雇主提供此類「模範員工」之資料訓練人工智慧，人工智慧經由學習後會察覺到「男性」要素之重要性，而給予正面評價，此會導致人工智慧日後之決策與建議偏好「男性」求職者⁵，而阻礙其他性別之錄取機會，此將有構成性別歧視之疑慮。典型的事例為亞馬遜公司於2014年使用人工智慧徵才，其後卻發現人工智慧系統因其所學習之資料而無法秉持性別中立原則，其給予女性求職者較低之評價，由於無法確保人工智慧系統將來不會作出歧視性之

排序，亞馬遜公司放棄此一人工智慧系統徵才計畫⁶。

二、企業可能導入人工智慧技術將徵才廣告自動投放於社群媒體，為達到廣告之最大效益，系統可依據企業之需求將求職者予以分類，並自動篩選決定投放廣告之目標族群。對於求職者而言，其能否看到廣告是經由人工智慧篩選，對於被排除之族群，即喪失應聘之機會。舉例言之，企業透過人工智慧於Facebook去投放招聘廣告，演算法僅會在目標族群的頁面上顯示廣告，故在女性用戶的頁面鮮少出現卡車司機的徵才廣告，而此類所謂「性別定位」(Gender Targeting) 投放廣告為社群媒體中常見之情況，此將使女性求職者無從知悉此一職缺之



3. Vgl. Hoeren/Sieber/Holzngel MultimediaR-Hdb/John Teil 29.4 Rn. 13.

4. Vgl. Waas, Künstliche Intelligenz und Arbeitsrecht, 2023, S. 118.

5. Dzida/Groh, Diskriminierung nach dem AGG beim Einsatz von Algorithmen im Bewerbungsverfahren, NJW 2018, 1917, 1918 f.

6. 參見周虹汶，亞馬遜AI徵才 歧視女性挨轟，自由時報新聞網站，2018年10月11日 <https://news.ltn.com.tw/news/world/paper/1238584> (最後瀏覽日：2024年1月22日)

存在，阻礙其進一步申請之可能性，對其就業有不利之影響，惟在此種情況下，並無「具體」、「個別」之求職者遭遇到不利之對待，現行法並無具體規範以處理此類「無受害者之歧視」(Opferlose Diskriminierung)⁷之問題，於大多數的情況下，潛在之受害者無法獲知其被排除而存在歧視之情事，自無從對企業求償，此亦是人工智慧普遍使用後可能產生之問題。

三、對於有意應徵工作職缺的求職者，有企業會採用「聊天機器人(Chatbots)」與求職者對話，導入人工智慧技術之聊天機器人具有深度學習之功能，從與使用者的應答中有能力延續對話之境況並予以回覆⁸。然而，若聊天機器人所提出之問題與「性別」相關之問題(例如懷孕)時，即產生在法律上應如何評價之問題，蓋人工智慧系統會依求職者之回答進行分類與評估，求職者若據實回答其真實情況(例如有懷孕計畫)，則可能被歸類到較無錄取希望之群體中，人工智慧之決策從結果上來看可能構成性別歧視。

四、部分企業可能會對於求職者之行為進行分析，其可能採取之方式包括自動文件分析、筆跡與語言分析、聊天機器人、自動工作面試，甚至是完整的人格特質分析。在企業實務中，尤以文本或語言分析較為常見⁹，透過人工智慧系統分析其語調、句子結構、使用詞彙等特徵，藉以評估求職者是否具有團隊協調能力、抗壓性、溝通能力、應變能力等特質¹⁰。演算法得就其分析之結果列出求職者之排名順序，以利企業作為選才之參考¹¹，當人工智慧系統對求職者之資料進行文本分析或語言分析時，亦有作成歧視性決策之風險。人工智慧系統對於有移民背景之求職者或非母語人士，可能會因使用語言的習慣、口音、方言而給予較為不利評斷，故可能會構成基於「種族」之間接歧視¹²。

參、性別歧視禁止之相關規範

我國憲法第7條為平等原則之規範，中華民國人民，無分男女、宗教、種族、階級、黨派，在法律上一律平等。此為平等原則之總則性規定，對於平等的要求除

7. Sesing/Tschech, AGG und KI-VO-Entwurf beim Einsatz von Künstlicher Intelligenz, MMR 2022, 24, 28.

8. Fink, Künstliche Intelligenz in der Personalarbeit, 2021, S. 70-71.

9. Hoeren/Sieber/Holzner/MultimediaR-Hdb/John Teil 29.4 Rn. 24.

10. Hoffmann, Möglichkeit und Zulässigkeit von Künstlicher Intelligenz und Algorithmen im Recruiting Personalsuche 4.0, NZA 2022, 19, 20.

11. Joos, Einsatz von künstlicher Intelligenz im Personalwesen unter Beachtung der DS-GVO und des BDSG, NZA 2020, 1216.

12. Dzida/Groh, Diskriminierung nach dem AGG beim Einsatz von Algorithmen im Bewerbungsverfahren, NJW 2018, 1917, 1919 f.

了消極面向之禁止恣意之差別待遇外，國家亦須積極採取平等保護之措施¹³。立法者在個別法規中定有「歧視禁止」之規範，就業服務法（以下簡稱就服法）第5條第1項設有禁止就業歧視之一般性規定，雇主對求職人或所僱用員工，不得以種族、階級、語言、思想、宗教、黨派、籍貫、出生地、性別、性傾向、年齡、婚姻、容貌、五官、身心障礙、星座、血型或以往工會會員身分為由，予以歧視。就服法第5條第1項屬於就業歧視禁止之原則性規範，若其他法律有明文規定者，從其規定。針對「性別、性傾向」之就業歧視問題，性別平等工作法（以下簡稱性工法）設有專章，在性質上應屬於就服法之特別法，而應優先適用。

若觀察我國就業歧視法制之發展歷程，早期職場上性別歧視之問題普遍較受到關注與討論¹⁴，2001年12月通過「兩性工作平等法」，其目的即在保障女性平等工作之權益，2008年更名為「性別工作平等法」，將多元性傾向者之工作權益納入保障，本法於2023年8月修正時更名為「性別平等工作法」，其立法目的在於消除職場中的性別與性傾向的歧視，確保性別工作權的平等，促進性別地位實質平等（性工法第1條規定參照）。性工法第7條至第11條明定雇主在招募、甄試、進用、分發、配置、考績、陞遷、教育訓練、



各項福利措施、薪資給付、退休、資遣、離職及解僱等方面，不得因受僱者之性別或性傾向而有差別待遇。為了落實就業歧視防治之目的，就服法第65條第1項、性工法第38條之1第1項皆針對違反者設有處罰規定，以行政制裁作為手段，敦促雇主確實踐行其法定義務。針對雇主民事損害賠償責任，依據性工法第26條之規定，受到性別或性傾向歧視之求職者、受僱者，就其財產上之損害，得向雇主請求損害賠償；若為非財產上之損害，求職者得依性工法第29條之規定向雇主請求賠償相當金額；其名譽被侵害者，並得請求回復名譽之適當處分。此外，性工法第31條設有舉證責任減輕之規定，求職者、受僱者僅須釋明差別待遇之事實，雇主即應就差別待遇之非性別、性傾向因素，或該受僱者或求職者所從事工作之特定性別因素，負舉證責任。

依據就服法第5條第1項規定，雇主對求職人或所僱用員工，不得以該法所禁

13. 參見李惠宗，憲法要義，第8版，2019年9月，頁148-149。

14. 關於立法背景之討論，參見焦興鎧，臺灣建構防制就業歧視法制之努力——二十年之回顧與前瞻，臺灣勞動法學會學報，第11期，2015年8月，頁4-6。

止之 18 項指標予以「歧視」。然而，就服法條文本本身並未針對「歧視」為進一步之立法定義。對照性工法之法條用語，第二章之標題雖訂為「性別歧視之禁止」，惟第 7 條至第 11 條之條文所使用之文字並非「歧視」而為「差別待遇」，如何理解「歧視」與「差別待遇」之意涵，此為首要釐清之問題。

對此，就服法第 5 條第 1 項之立法理由或可提供一線索，其指出「歧視係指雇主無正當理由而恣意實施差別待遇而言」，由此觀之，「歧視」在意義上似非完全等同於「差別待遇」。若對照性工法之條文用語及架構，求職者或受僱者縱因「性別」或「性傾向」而在招募、甄試、進用、分發、配置、考績或陞遷上遭受雇主之「差別待遇」，但若其符合性工法第 7 條但書「但工作性質僅適合特定性別者」之要件，其差別待遇則可被認為具有「正當事由」，因而被評價為合法。換言之，唯有差別待遇被評價為違法時，始能以「歧視」稱之。由此觀之，「差別待遇」一詞應有較為客觀、中性之涵義，其表述一種求職者或受僱者遭到「不利對待」之事實狀態，而「歧視」則有某種程度之法律評價¹⁵，或者可以說，「歧視」應為無正當理由（不合法）之差別待遇。因此於雇主是否違反「歧視禁止義務」之判斷上，應檢視雇主是否基

於「性別、性傾向」而給予求職者或受僱者差別待遇，若有差別待遇之存在，進一步則須檢視其是否具備「正當事由」，茲分述如下：

一、直接或間接之不利對待

關於「差別待遇」之定義與種類，參照性工法施行細則第 2 條之規定，其係指雇主因性別或性傾向因素而對受僱者或求職者為「直接或間接不利之對待」。所謂「直接之不利對待」，係指以歧視指標作為差別待遇之事由，對求職者或受僱者為直接不利之對待。舉例言之，雇主於為某項約定或為措施時，使之僅對某一性別之求職者或受僱者發生效力，其構成基於性別直接之不利對待¹⁶。典型之事例為雇主在招聘的廣告中明白表示該招聘的職位「限女性」或「限男性」，此為基於性別而對其他性別之求職者為不利之對待，在判斷上較無疑問。



15. 關於「歧視」、「差別待遇」在我國不同法規使用之分析與意涵，詳見黃夢萱，就業歧視禁止之界限——論就業歧視禁止項目與差別待遇之正當理由，臺灣大學法律學院法律學系碩士論文，2016年1月，頁13-20。

16. 參見黃程貴主編，個別勞動法，2021年3月，頁346（林佳和執筆）；林更盛，論性別歧視與母性歧視之禁止——評台灣高等法院八十七年度勞上易字第一號判決，月旦法學雜誌，第74期，2001年7月，頁187。

惟若雇主在形式上並未明確表示以歧視指標「性別」作為差別待遇的理由，而是用外觀上中立的標準、約定或措施為區分之依據，然此等中性的標準事實上會讓「具特定歧視指標（如某一性別）之人」受有負面之影響，此將構成「間接之不利對待」¹⁷。舉例言之，若某項標準、措施從外觀看來是對於部分工時受僱者較為不利，然若依實際的統計數字可得知，大多數的從事部分工時工作者多為女性，在此情況下則有可能會構成基於「性別」而生之間接不利對待¹⁸。又若雇主在形式上雖無明確表明以性別，而是透過其他標準（如身高、體重）作為區分的依據，若事實上因該約定或措施將使某一性別之勞工遭受不利待遇明顯較另一性別多，則亦構成因性別而生的間接之不利對待¹⁹。

當企業使用人工智慧從事招聘工作，透過演算法在過程中直接代替雇主剔除不合適人選而自行做成決策，此是否構成不利之「對待」？若考量到「對待」之文義，

其似應指涉人類之行為，無論如何應由人類參與其中²⁰，在此理解下，則可能得出由人工智慧所作成之建議或決定無法被評價為不利對待之結論。對此問題，拙見以為，考量到性工法歧視禁止規定之目的在於保障工作權之性別平等、貫徹憲法消除性別歧視、促進性別地位實質平等之精神²¹，不宜直接將人工智慧之決策一律排除於「對待」之文義之外，否則將使立法目的落空，故應對本法之「對待」採取較為廣義之解釋，只要雇主將其人事決策之選擇空間委由人工智慧進行，亦應認為此屬於歧視禁止法制所規範之「對待」²²。

至於如何認定「不利」對待，參酌文獻之討論，此應由客觀上一般理性第三人的觀點作為判斷標準²³。在招聘程序中，雇主最終所作成拒絕錄用之決定固然屬於不利之對待，當求職者在前一階段即被排除而未獲得面試機會，解釋上亦應為不利對待²⁴，此不因為雇主使用人工智慧篩選面試者而有不同之評價。以人工智慧

17. 參見黃程貫主編，個別勞動法，2021年3月，頁346（林佳和執筆）；傅柏翔，美國與臺灣就業歧視法制之舉證責任與適當比較基礎問題初探，法令月刊，第69卷第9期，2018年9月，頁76；鄭津津，就業上年齡歧視之探討，月旦法學雜誌，第184期，2010年9月，頁179。

18. BeckOK ArbR/Roloff AGG § 3 Rn. 17.

19. 林更盛，論性別歧視與母性歧視之禁止——評台灣高等法院八十七年度勞上易字第1號判決，月旦法學雜誌，第74期，2001年7月，頁187-188。

20. Vgl. Hans Steege, Algorithmenbasierte Diskriminierung durch Einsatz von Künstlicher Intelligenz – Rechtsvergleichende Überlegungen und relevante Einsatzgebiete, Multimedia und Recht (2019), S. 715/718.

21. 參見性工法第1條第1項之規定。

22. 參見洪瑩容，企業使用人工智慧之聘僱歧視問題研究——兼論對於現行反歧視法制之檢討，興大法學，第34期，2023年11月，頁23-25。

23. Vgl. BAG, Urteil vom 23. 8. 2012 – 8 AZR 285/11, NZA 2013, 37.

24. Freyler, Robot-Recruiting, Künstliche Intelligenz und das Antidiskriminierungsrecht, NZA 2020, 284, 287; von Lewinski/ de Barros Fritz, Arbeitgeberhaftung nach dem AGG infolge des Einsatzes von Algorithmen bei Personalentscheidungen, NZA 2018, 620, 621.

系統處理求職者資訊時，演算法會將資訊依關聯性予以分類，而如果人工智慧進一步被運用於決定邀請面試甚至是聘用之人選時，其自然依此分類篩選人才，當人工智慧系統之分類基準涉及到「性別」此一法律所禁止之指標時，即可能違反性工法歧視禁止之規定。以人工智慧所進行之分類與決策而言，其通常不會取決與單一之「指標」，故幾乎不會發生直接不利對待之情況。對於求職者，其面臨者較為「間接之不利對待」，蓋演算法之基礎在於資料探勘（Data Mining），利用龐大的資料庫去建立模型，從中找出隱藏的特殊關連性，並將求職者進行分類、評估，在運行的過程中，演算法不僅可能在「未被法律禁止之指標」中找出其關聯性，亦可能觸及到「法律所禁止之指標」。當演算法所連結者為一表面上看起來中性之指標而對求職者進行分類與評估，惟其結果卻使性工法所欲保護之群體（例如女性）遭受不利對待之比例較為顯著，此即可能產生間接不利對待之風險²⁵。誠如前文所提及之例子，當企業為使人工智慧建立出「職業成功之預測模型」，其若是以過去模範員工之資料作為訓練學習之資料，即便所收集之資料僅反映出過去受僱個人及群體之績效，但由於過去女性受僱之比例較低，或者是其在企業擔任主管職之比例較低，

人工智慧系統即可能會辨識出「男性」與成功、績效間之關聯，其將可能導致女性求職者在分類與決策上遭受較為不利之評價與對待。

二、正當事由

從性工法第7條但書之規定可知，只要「工作性質僅是適合於特定性別者」，雇主仍得因為「性別」或「性傾向」而為差別待遇。換言之，只要雇主具備正當事由，其所為之差別待遇並不會被評價為違法之歧視。至於如何理解性別工作平等法第7條但書所謂「其工作性質僅適合特定性別者」？同法施行細則第3條進一步將但書之規定解釋為「非由特定性別之求職者或受僱者從事，不能完成或難以完成之工作」。意即該工作有取決於生理性別上的必要性，或者法律對於特定性別設有禁止僱用之規定²⁶，則雇主不僱用女性應有正當事由。若雇主委由人工智慧進行資料分析及篩選時，對於雇主而言，由於人工智慧之決策具有不可預測性，對於雇主而言，亦可能難以證明其具有正當事由，蓋演算法透過資料分析僅能找出性別與某工作特性之關聯，例如某一種性別在特定職位上的表現比另一性別更成功，但其並無法確認其是否具有因果關係，亦即特定性別是適合特定工作之原因²⁷。即便如此，當企業在聘僱程序中使用人工智慧，求職

25. Vgl. von Lewinski/ de Barros Fritz, Arbeitgeberhaftung nach dem AGG infolge des Einsatzes von Algorithmen bei Personalentscheidungen, NZA 2018, 620, 622; Waas, Künstliche Intelligenz und Arbeitsrecht, 2023, S. 127.

26. 參見洪瑩容，非特定性別不能履行之工作——以德國法制之觀察為中心，月旦法學，第273期，2018年2月，頁203-206。

者要能成功主張受到就業性別歧視仍非易事，此涉及到舉證責任之問題，將於以下說明之。

三、舉證責任

當企業使用人工智慧進行招聘決策時，對於求職者而言，另一個可能遭受到之難題在於尋求權利救濟時之舉證困難。性工法第31條雖設有求職者、受僱者舉證責任減輕之規定，當求職者「釋明」差別待遇的事實後，雇主應就差別待遇之非基於性別、性傾向因素或該求職者所從事工作之特定性別因素負舉證責任。然而，相較於傳統之招聘方式，求職者與真實人類之互動可感知到不友善之反應、情緒，由人工智慧系統所主導之聘僱程序，求職者通常欠缺此種可被感知之信號，而難以察覺受到「歧視」。此外，由於人工智慧系統存在自主決策之不透明，演算法之黑箱將讓使用者難以在事後確定人工智慧之決策原因²⁸。且演算法之相關資訊亦可能因事涉營業秘密²⁹而不被公開，此皆使因果關係之證明陷入困境³⁰。職此，在人工智慧蓬勃發展之同時，應思考如何透過對於人工智慧系統使用之管制或者舉證責任之調整，解決就業歧視之問題，減輕受歧視之求職者在權利主張上之障礙。



► 肆、違反性別歧視禁止義務之法律效果

若求職者於聘僱程序中遭受性別歧視，可向地方主管機關進行申訴，雇主一旦被認定有就業歧視之情事則可能遭受主管機關處以罰鍰³¹。除了行政罰外，違反歧視禁止規定之雇主亦須對求職者負損害賠償責任。當企業使用人工智慧系統進行招聘工作，若其因此造成性別歧視之結果，雇主即可能須對求職者依性工法第26條之規定負損害賠償責任。問題在於，若人工智慧系統在未經「人類」之介入而作成具有性別歧視之決策，此時雇主是否須為人工智慧系統之自我決策負責？此問題涉及到性工法第26條之責任性質，亦即該條規定究竟為「過失責任」或「無過失責任」之爭議。拙見以為，性工法第26條之損害賠償責任仍應為「過失責任」之規

27. Freyler, Robot-Recruiting, Künstliche Intelligenz und das Antidiskriminierungsrecht, NZA 2020, 284, 287.

28. 關於人工智慧之自主性風險，參見呂彥彬，人工智慧時代契約法之挑戰與回應，政大法學評論，第169期，2022年6月，頁305。

29. Freyler, Robot-Recruiting, Künstliche Intelligenz und das Antidiskriminierungsrecht, NZA 2020, 284, 290.

30. 相關討論請參見洪瑩容，企業使用人工智慧之聘僱歧視問題研究——兼論對於現行反歧視法制之檢討，興大法學，第34期，2023年11月，頁26-33。

31. 參見就服法第65條、性工法第38-1條之規定。

定³²，換言之，性工法第26條之適用，應以雇主違反歧視禁止義務具有可歸責事由時，始對求職者負損害賠償責任。在此見解下，接續要處理之問題為，雇主得否以其無從預見人工智慧決策之過程，進而主張自己就該歧視之決策欠缺可歸責事由而不負賠償責任？在此涉及到雇主應否為人工智慧之歧視決策負責之問題，而核心的爭議在於，人工智慧可否被評價為民法第224條之債務履行輔助人？對此問題，拙見以為，雇主對於人工智慧所為具有歧視性之自我決策，在現行法下似難有類推適用民法第224條規定之空間，其主要之理由在於欠缺「可資比較之利益狀態」，由於民法第224條在性質上應屬於一歸屬之規定，須以履行輔助人本身具有「故意過失」之「行為」為其前提，人工智慧之自主決策非「人類」意思所支配下之「行為」，自難成為被歸屬之客體，故欠缺類推適用之基礎。惟此一結果難符事理之平，正本清源之道仍需透過立法途徑解決³³。

► 伍、結語

在現代的勞動世界中，人工智慧的採用已然成為現實，尤其用於企業招聘工作中，以目前的技術而言，尚無法完全克服演算法偏見之問題，因而可能產生聘僱歧視之問題。在現行法制下，由於人工智慧系統之自主決定性格以及不透明性，如何能察覺並進一步認定人工智慧所為之決策構成性別歧視，實屬不易，此可能是現行反歧視法制遭遇到適用上之困難。在立法上或可從事前對於人工智慧之管制，以及要求人工智慧之使用者（雇主）在使用人工智慧前應負有測試義務，並將此等測試過程予以紀錄，將來在訴訟中若雇主無法提出相關紀錄，則可認為勞工已盡釋明責任，雇主即須證明其所為之差別待遇具備正當事由。另一個重要之問題為：當人工智慧所作成具有歧視結果之決策時，採用人工智慧之雇主應負何等責任？於現行法尚存可否適用民法第224條之疑義下，吾人應思考將來如何透過立法以妥適處理人工智慧使用者責任之問題。

32. 詳細論述請參見洪瑩容，企業使用人工智慧之聘僱歧視問題研究——兼論對於現行反歧視法制之檢討，興大法學，第34期，2023年11月，頁42-47。

33. 詳細論述請參見洪瑩容，企業使用人工智慧之聘僱歧視問題研究——兼論對於現行反歧視法制之檢討，興大法學，第34期，2023年11月，頁47-51。