國立臺灣大學法律學院科際整合法律學研究所

碩士論文

Graduate Institute of Interdisciplinary Legal Studies

College of Law

National Taiwan University

Master's Thesis

人工智慧金融與金融歧視之研究——

以信用貸款為中心

Research on Artificial Intelligence Finance in Financial Discrimination:

Focusing on Credit Lending.

吳媚烜

Mei-Hsuan Wu

指導教授：楊岳平 博士

Advisor: Yueh-Ping Yang, S.J.D.

中華民國 114 年 1 月

January 2025

doi:10.6342/NTU202500326

國立臺灣大學碩士學位論文
口試委員會審定書

人工智慧金融與金融歧視之研究—
以信用貸款為中心
Research on Artificial Intelligence Finance in Financial Discrimination:
Focusing on Credit Lending.

本論文係吳媚烜君（學號 R09A41020）在國立臺灣大學科際整合法律學研究所完成之碩士學位論文，於民國 114 年 1 月 23 日承下列考試委員審查通過及口試及格，特此證明

指導教授：

口試委員：

doi:10.6342/NTU202500326

謝辭

又到了寫下謝辭的時刻，意味著我的第二個碩士生涯即將劃下句點。回憶起在科法所學習的時日，是一段難忘且深刻影響著人生規劃的歲月。我的音樂背景總讓我在法律人中顯得格外不同，也因為毫無任何學科知識背景的支持，習法之路是格外辛苦的，所幸在這段日子裡，身邊總有許多優秀且溫暖的人們鼓勵、幫助著我，讓我順利完成學業，也順利於在學中通過律師考試成為一名正式執業律師。

首先感謝指導教授及導師楊岳平教授的教導及建議，從碩一開始跟隨著老師的每一堂課程、演講以及的導生活動中，老師的智慧及人生態度是我心中仰望的楷模，論文也因為有老師非常仔細的指導和審閱才得以成形，也非常感謝口試委員清華大學林勤富教授及東吳大學林育廷教授給予的寶貴建議，使我得以補充、思考不足之處。

感謝同為楊門子弟的鎧豪，從碩一開始到口試的每一個重要的日子裡，都有你的協助及令人安心的陪伴，很幸運也很開心能一起從老師的導生進階到指導學生，也很感謝你在每一個受挫的時刻，都為我指點迷津、給予許多真誠的建議；也感謝同門的哲綸，在導生聚熟識後，在簽指導的過程中很感謝學長許多寶貴的建議，也感謝學長總是熱心召集新年拜年及教師節活動，讓大家能在重要的時刻裡與老師一同團聚。

感謝 R09 同學們的同窗情誼，育雯是我在臺大第一位認識的同學，同為臺藝大校友的我們在初次見面就有著熟悉感，一起經歷這幾年無數的歡笑及淚水，跟你在一起學習或是出遊的相處總令人覺得很舒適及默契，是很難得的夥伴；育雯、佑真、怡婷、舒涵是很早就一同修課、吃飯、玩耍的夥伴，一同經過了許多風雨，也一同創造了許多回憶；東碩、立達、育誠和廷韋是宅團夥伴，在每一次的飯局、出遊或是教室深夜談話中，有你們的參與總是帶來許多的歡笑；立達、羽希、日弘是登山兼讀書會好夥伴，感謝你們一起的合歡山行，讓我得以嘗試第一次的百岳行，也感謝一起參與

doi:10.6342/NTU202500326

讀書會的討論，讓我擁有互相砥礪的戰友；二哥是唯一在科法所可以一同聊育兒經的好夥伴，也只有同需兼顧幼兒家庭生活和課業的戰友，才能深知箇中滋味。

感謝院辦同事們，院辦工作讓我的研究所生活增添幾分色彩，也因此認識許多好友。顏平是科法所的直屬，而我們卻是在院辦工作才開始熟識，總是能從你身上得到許多新知，非常感謝你的陪伴和鼓勵，與你的每一次跑步、出遊、吃飯和談天，都是非常愉快且收穫滿滿的；鐙翔是院辦第一天上班就認識的夥伴，熱情、聰明的你在第一次工作中就一見如故，有你一起工作的日子都很愉快，也很感謝你在我準備國考期間的鼓勵和協助；婷瑄是在離開院辦工作後還持續有聯絡的夥伴，很佩服你對空瑜的熱情，熟識後也覺得你十分親切可愛且相處起來很舒服；書緯、雲珽、明寬是我在院辦後期工作的小夥伴，總是青春、有活力地帶來許多歡笑，也感謝總能和我一起討論許多國考或是論文的問題。

律訓第 32 期第一梯次的第六組可愛的夥伴們，奕賢、昆鴻、顥源、天懷、穎昌、馨萱、宥睿、沐恩、庭裕，與你們相處的日子就像是再青春了一次，總是溫暖地在每一個重要的時刻陪伴著彼此，一起分享工作上或是生活上的喜與憂，互相鼓勵彼此順利度過每一個重要的階段，感謝這份難得的情誼。

能夠無後顧之憂地追尋自己的理想，歸功於家人們的支持，最感謝的是 Alan 和炫錕，陪伴我經歷著低潮、挫折，忍受我的焦慮和脾氣，也與我一同享受學習的富足和成果，未來的日子我們都會越來越好，一起勇敢地探索這個世界吧！最後，我想好好感謝自己，轉換跑道的日子裡嚐盡了多少苦只有自己最清楚，感謝自己能夠不被各種質疑的聲音所打敗，堅持著自己相信的理想和目標，未來的路肯定仍充滿各種挑戰，期許自己仍繼續不忘初心、保有熱情地迎接所有未知，成為更好的自己。

吳媚烜 謹誌

2025 年 2 月 1 日於臺北

doi:10.6342/NTU202500326

摘要

隨著人工智慧技術的發展，其於各行業中的應用日益普及，金融機構也不例外地在金融服務中開始運用人工智慧技術，藉此提升效率以提供更完善的金融服務。但於此同時，人工智慧技術的運用也帶來了公平性的隱憂。

本文以金融機構運用人工智慧技術於信貸服務中引發的歧視問題為討論主軸，從目前實務上發生的案例，發現人工智慧信貸不同於傳統信貸的歧視風險態樣，進而梳理人工智慧信貸歧視的風險形成與原因，並指出我國現行金融反歧視法存在規範空白的問題。本文進而透過比較法研究討論實務和學說如何在現行金融歧視法規範的架構下解決人工智慧信貸歧視問題，並以美國公平信貸相關規範及學者研究文獻為主要參考對象，觀察傳統公平信貸法制實踐於人工智慧信貸時產生的公平性問題與解釋挑戰。

透過比較法的研究，本文歸納我國法制針對人工智慧信貸應用可考慮的監理路徑，並以《金融業運用人工智慧（AI）指引》及《金融機構運用人工智慧技術作業規範》的規範內容為主，提出我國金融監理面對人工智慧信貸歧視問題時可考慮的具體規範建議。

關鍵詞：金融反歧視法、公平信貸、人工智慧信貸歧視、機器學習、金融消費者保護

doi:10.6342/NTU202500326

ABSTRACT

With the development of Artificial Intelligence (AI) technology, its application in various industries is becoming more and more popular. Financial institutions are no exception, which have started to utilize AI technology in financial services to enhance efficiency and provide better financial services. However, at the same time, the use of AI technology in financial services also brings about the hidden problem of fairness.

This thesis focuses on the use of AI technology in credit services by financial institutions. It identifies the different patterns of discrimination of AI credit services from that of traditional credit services from the cases occurring in practice, explores the causes of AI credit discrimination, and specifies the gaps in the existing financial anti-discrimination laws in Taiwan. Furthermore, the thesis adopts the comparative legal study to discuss how the practice and theory can solve the problem of AI credit discrimination under the existing financial anti-discrimination legal framework in other countries. It mainly focuses on the legal reforms in the U.S. related to fair credit and scholars' literatures to observe the interpretation challenges related to the fairness problem when applying the traditional fair credit laws to AI credit services.

This thesis finally discusses how the current legal system can be applied in the context Taiwan's law, specifically the Guidelines on the Use of Artificial Intelligence (AI) in the Financial Sector and the Code of Practice on the Use of Artificial Intelligence by Financial Institutions. It proposes specific regulatory suggestions to address the problem of discrimination in the use of AI in credit services.

KEYWORDS: *Financial Anti-Discrimination Act, Fair Lending, Artificial Intelligence Credit Discrimination, Machine Learning, Financial Consumer Protection*

doi:10.6342/NTU202500326

目次

doi:10.6342/NTU202500326

doi:10.6342/NTU202500326

圖次

doi:10.6342/NTU202500326

第一章 緒論

第一節　研究動機

身處數位時代，金融科技的發展正顛覆著傳統金融服務的模式，隨著大數據分析、機器學習（machine learning）和人工智慧技術（artificial intelligence）的進步，人工智慧系統開始應用於金融服務中，提高作業效率、降低成本，同時可能得到更加精確、客觀的風險評估[1]，依據金管會 2024 年之統計資料顯示，目前我國金融機構及金融周邊單位中，已有 108 家於金融服務中應用人工智慧工具，銀行業在其中的比例為 4%、壽險業為 62%、產險業為 50%、證券期貨業則約為 18%[2]，該資料中也統計金融業使用人工智慧及聊天機器人之使用情形，主要用途用於內部行政作業，使用比例達 20%，用於智能客服之用途的使用比例則約 16%，其他用途則像是行銷廣告、認識客戶風險以及防範金融犯罪等[3]。

人工智慧技術所帶來的優點顯而易見，但風險也日益顯現，特別是關於公平性原則的隱憂。反歧視是平等原則的體現，是現代社會運行的重要原則也是正義與人權保障的具體落實，對於金融服務而言，公平性將直接影響到經濟上的平等與穩定，因此公平性原則的落實與弱勢族群的權益保障顯得格外重要。特別在信用貸款的服務，其對於個人消費者或是企業而言為重要資金獲取途徑，在申請貸款、評估信用分數及決

[1] 陳凱迪（2018），〈人工智慧發展對金融業之衝擊與因應〉，《財金資訊季刊》，93 期，頁 16-19；金融監督管理委員會（2023），〈金融業運用人工智慧(AI)之核心原則與相關推動政策〉，頁 2。

[2] 此為 2024 年期貨公會舉行之台北國際期貨論壇演講，演講中之簡報所表示之統計數字，另可參前揭註 1 金管會公告之〈金融業運用人工智慧(AI)之核心原則與相關推動政策〉頁 10，其顯示 2023 年 5 月之調查結果，我國 175 家金融機構中有 63 家採行人工智慧技術，其所使用的用途包含客群經營、風險管理及法令遵循、流程精進、數據分析及其他用途。

[3] 陳林幸虹，〈台灣金融業使用 AI 占比近 3 成金管會發布指引〉（06/20/2024），《中央廣播電臺》，https://www.rti.org.tw/news/view/id/2210313。

doi:10.6342/NTU202500326

定是否核貸的流程中，所有申請者都應受到公平的對待，且應有完整、全面的法律規範作為保障[4]。

人工智慧運用的普及產生了許多倫理問題，信貸服務中亦因此產生了許多公平性問題，其可能重現並加劇社會既有的不平等現象[5]。首先，演算法的設計與運作係基於大量的資料輸入，但資料可能帶有歷史上偏見與不平等的風險，例如過去的歧視現象所產生的歷史偏見會反映在資料中，進而影響演算法的決策。此外，演算法設計中的變項選擇和權重設定也可能產生偏見，例如過分仰賴特定變項，如薪資收入、教育背景等，可能導致對某些群體的不公平對待。不當的演算法設計或運行，可能來自於人為的故意，亦可能是無意的數據偏見或設計缺陷，其將對特定群體產生不公平對待的歧視結果，固化既存的偏見，甚至產生基於數據資料而難以察覺的隱性偏見[6]。

目前國際組織與美國、歐盟所訂定的人工智慧規範中，皆可見強調「以人為本」的公平、反歧視之準則，例如 2021 年聯合國教科文組織（UNESCO）公布了「對 AI 的倫理建議（Recommendations on the Ethics of AI）」中即強調「公平無歧視」原則[7]。在人工智慧快速發展並應用於金融服務之際，面對日新月異的科技進步而帶來的新議題，實有必要透過立法防止歧視並落實公平原則，以及從監管的面向補足法律規範不足之處，有效、持續地維護公平。

在信貸申請的流程中，為了評估借款人的信用分數，金融機構多半透過資產、個人能力（如學歷、職業）與個人條件（如年齡、性別、健康）作為信用評分的核心基礎[8]，透過建立不同的統計模型，將個人條件作為變項進行計算，作為決定是否核貸的

[4] Jason Jia-Xi Wu, *Algorithmic Fairness in Consumer Credit Underwriting: Towards a "Harm-Based" Framework for AI Fair Lending*, 21 BERKELEY BUSINESS LAW JOURNAL, 69 (2024), https://doi.org/10.15779/Z38CF9J785 (last visited Jan 3, 2025).
[5] Rikka Koulu, *Human Control over Automation: EU Policy and AI Ethics*, EUR. J. LEG. STUD. 9, 11 (2020).
[6] Joshua A Kroll et al., *Accountable Algorithms*, 165 UNIV. PA. LAW REV., 675 (2017).
[7] UNESCO, *Recommendation on the Ethics of Artificial Intelligence*, 11 (2021) https://unesdoc.unesco.org/ark:/48223/pf0000380455 (last visited Jun 11, 2024).
[8] Josh Lauer, *Creditworthy*, COLUMBIA UNIVERSITY PRESS, 8 (2017).

doi:10.6342/NTU202500326

重要工具。在人工智慧和大數據時代下，金融機構能夠更廣泛蒐集借款人的各種資訊，也透過機器學習加速並減輕計算信用分數的業務，提供更高效率、更低成本的授信服務[9]。

然而，更多看似無害的個人資訊的蒐集，在應用機器學習以評估信用時，卻導致了隱性的歧視產生。除了傳統反歧視法認定的受保護特徵之外（如年齡、性別、種族...等），個人居住地的郵遞區號、政治傾向、心理測驗表現、線上支付的使用紀錄、社交平台的使用.....等，看似中性的日常生活足跡特徵作為個人資料之一部分，在演算法的運作中對某些特定特徵有不當的權重比例，進而產生帶有偏見的結果，進一步導致隱性的歧視發生[10]。

我國目前尚未見得因人工智慧系統應用所導致的信貸歧視案例，從前揭金管會統計的資料也發現目前銀行業應用人工智慧系統的比例僅有 4%，且其用途多用於內部行政作業，然而根據歐洲議會和理事會 2024 年 6 月 13 日頒布的第 2024/1689 號條例，第 58 段引言提及：「用於評估自然人信用評估的人工智慧系統應被歸類為高風險人工智慧系統，因為它們決定了這些人獲得金融資源或住房、電力和電信服務等基本服務的機會。用於此目的的人工智慧系統可能會導致對個人或群體的歧視，並延續歷史上的歧視模式，例如基於種族或族裔、性別、殘疾、年齡、性取向的歧視，或產生新形式的歧視性影響。」依此說明，信貸人工智慧系統將被歸類為高風險人工智慧而受有相關規範；該法案第 6 條第 2 項附件 3 列舉高風險人工智慧系統的項目中，第 5 條〈獲得和享受基本私人服務和基本公共服務及福利〉之第 b 款提及：「用於評估自然人信用或建立其信用評分的人工智慧系統，但用於檢測金融詐欺目的的人工智慧系統除外。」由此見得，用於信用評分的人工智慧系統屬高風險人工智慧系統，有更高的義

[9] FINANCIAL STABILITY BOARD (2022), FINTECH AND MARKET STRUCTURE IN THE COVID-19 PANDEMIC, https://www.fsb.org/wp-content/uploads/P210322.pdf (last visited Jun 11, 2024).
[10] Katja Langenbucher, *Consumer Credit in The Age of AI-Beyond Anti-Discrimination Law*, LAW WORK. PAP., 9 (2023).

doi:10.6342/NTU202500326

務必須遵守，故縱使我國金融實務上似無規範人工智慧信貸歧視的急迫性，但仍應重視並完善相關規範，且在我國金融業的監管密度嚴格的現況下，透過預先完備規範與相應配套的監管措施，也才有機會在金融實務上實際應用人工智慧工具，以協助並提升金融服務的品質及效率。

因此，本研究之目的在於研究我國現行的法律是否足以規範上述的人工智慧信貸歧視現象？若無法有效防免，應如何從規範面向上著手改善？期待可為我國未來的政策制度、立法研擬提供研究成果及建議，並從事前、事後的角度介入，討論監管的職責、確定相關的責任主體，建立得以持續監管之機制，確保人工智慧的貸款決策符合反歧視法律的要求，防範不同型態的歧視風險。

第二節　研究對象與範圍

演算法的定義是「在有限步驟內解決數學問題的程序」[11]，亦即，演算法是一組包含問題輸入及結果輸出的一組運算規則，其用途通常是為了資料處理、計算或推理，用以解決問題而設計的運算規則[12]。在人工智慧的應用中，為了讓電腦模仿人類的認知，針對特定問題做出決策、判斷結果，必須仰賴機器學習訓練演算法，藉此分析資料並據此做出決策的演算法使用機器學習技術來處理資訊並自動生成評估或決策，而無需直接人工干預[13]，盡可能作出最佳的決策。

然而，演算法並不僅限於人工智慧領域之應用，並非所有的演算法都與人工智慧相關， 演算法也應用於許多其他領域中，例如密碼學、數據分析、物理學、遊戲等領域[14]。而本文的主要研究對象，為人工智慧運用於金融領域中之信用貸款服務時所產

[11] Merriam-Wester.Com, *Algorithm*, https://www.merriam-webster.com/dictionary/algorithm.
[12] 謝孫源、李佳衛、洪綾珠，〈電機資工的現況與未來：生活中的演算法〉，《科技大觀園》，https://scitechvista.nat.gov.tw/Article/c000003/detail?ID=084a957b-aa02-4833-a08d-ec0788e97d50（最後瀏覽日 12/11/2024）。
[13] Datenschutz-notizen, *Privacy and AI: Schufa algorithm condemned by the CJEU*, https://www.datenschutz-notizen.de/privacy-and-ai-schufa-algorithm-condemned-by-the-cjeu-0946847/ (last visited Nov 10, 2024).
[14] Forbes, *What Is An Algorithm? Defining And Applying Algorithms*, https://www.forbes.com/sites/technology/article/what-is-an-algorithm/ (last visited Jan 05, 2025).

doi:10.6342/NTU202500326

生的歧視與公平性問題，因此本文所稱的人工智慧系統，係指透過資料學習、機器學習或相關建立模型之演算法，進而進行決策之系統[15]，其中行文所提及的演算法僅限於人工智慧及機器學習運作程序中所使用的演算法，非廣義的演算法定義，為了行文的流暢性，文中可能會有「演算法」、「人工智慧」、「機器學習」等用詞出現，皆係指限縮於指涉應用人工智慧及機器學習技術之程序，尚不論及此範圍外的意義，先予敘明。

本文旨在透過應用人工智慧系統的信用貸款為主要研究對象，研究人工智慧應用人工智慧系統應用於信用評估的金融歧視問題，因此，本文涉及的人工智慧使用範圍，在於金融機構利用人工智慧工具協助信用評估，透過其評估的結果做出最終的貸款決策。另外，考量目前人工智慧金融運用的實際案例情形，本文除了信用貸款外，也將涉及信用卡的信用評估納入本文之研究範圍，以囊括涉及用以信用評估的人工智慧金融歧視問題。

第三節　研究方法與限制

第一項　文獻回顧法

為研究人工智慧信貸歧視的成因、類型與解決方法，本文將針對歧視的定義與分類、反歧視法之規範方式、監理機關之持續監管及我國立法改善之處等面向，蒐集彙整與人工智慧歧視或信貸歧視相關之國內外期刊論文、政策性文件及新聞媒體資料，透過分析現有之文獻資料，釐清問題目前的研究成果與理論基礎。

第二項　比較法研究

我國法目前上缺乏比較法上的公平信貸規範，僅有針對金融服務應用人工智慧技術頒布建議性的指引及自律規範。故本文將從涉及信貸的公平信貸規範及人工智慧技

[15] 金管會（2024），〈金融業運用人工智慧(AI)指引〉，頁 3。

doi:10.6342/NTU202500326

術的公平性風險兩個面向，將美國法作為主要比較對象，特別是參酌美國公平信貸相關規範、人工智慧信貸的學術文章以及政府部門的文件或討論，尋求面對人工智慧信貸新挑戰的可能解決路徑，礙於篇幅及研究範圍，本文在比較法研究上僅針對與研究主題相關的規範節錄說明、分析，並不會針對法規範整體有完整的論述。

第四節　研究架構

本論文共分六個章節，從信用貸款業務中的徵信與授信業務為主軸，探討人工智慧系統應用於其中之審核與決策時可能產生的歧視風險。進而探討金融反歧視之核心目標、立法制度，再將其應用至人工智慧信貸歧視時應如何避免風險的產生。最終透過比較法研究，從我國法的現行體制下，提出適當的立法芻議與監管措施。

第一章為〈緒論〉，確立本文聚焦於人工智慧系統應用於金融服務、信用貸款之研究對象及範圍，特別在公平性與歧視的疑慮，並以文獻回顧法及美國法為主之比較法研究作為研究方法。

第二章為〈人工智慧信貸的歧視問題〉，本章先定義何謂人工智慧歧視、人工智慧信貸歧視，從人工智慧的運作原理與公平性風險談起，並觀察實務上發生的人工智慧信貸歧視案例以提出解決此問題的必要性，進而探討歧視之定義、形成及產生的原因。

第三章為〈公平信貸規範的理論與法制〉，本章主軸在於傳統公平信貸規範的討論，借鑑比較法的規範，特別是美國平等信用機會法及相關法令，分析公平信貸法規範的目的及法定之受保護特徵，再從美國法與歐盟法談差別對待歧視（直接歧視）與差別影響歧視（間接歧視）。

第四章為〈人工智慧信貸歧視之解決之道〉，本章主要以美國進行比較法研究，探討現行法制在人工智慧信貸歧視問題所面臨的挑戰，從公平信貸規範出發，討論法

doi:10.6342/NTU202500326

律制度的應用、監管機關的監管方式，並試圖討論較小歧視方案及可解釋人工智慧用以解決人工智慧信貸歧視的可能性。

第五章為〈我國法下之人工智慧與公平信貸規範〉，從我國金融法的現行規範談起，試圖以《金融業運用人工智慧（AI）指引》及《金融機構運用人工智慧技術作業規範》做為規範依據，並參酌比較法的規範建議為其提出建議，從立法上建構基本的定義，依我國現況提出適當的立法及監管建議。

第六章為〈結論〉，總結本論文針對人工智慧系統於信用貸款中的歧視討論。本文參酌美國法針對信用借貸的公平性問題規範後，借鑑其立法及監管模式，結合可解釋人工智慧的技術考量，回顧我國現行金融反歧視法規的規範不完整，進而提出我國可能解決人工智慧信貸歧視的立法及監理芻議。

doi:10.6342/NTU202500326

第二章 人工智慧信貸的歧視問題

過去在信用貸款的業務中，金融機構須倚靠人力個案評估，舉凡信用評分、風險評估、授信決定等流程。在信用評分中，個人條件影響信用分數的高低，為審核的關鍵因素。在人工智慧迅速發展之際，金融業嘗試在信貸程序中引進人工智慧系統，加速審核流程以提高效率、節約人力需求，但也因此產生或加劇金融領域的風險，引起金融消費者和投資者的疑慮，如存在偏見、不公平或歧視性消費者的結果，或資料管理、使用及隱私安全方面的擔憂[16]。本章先梳理人工智慧歧視的公平性問題，再討論目前人工智慧信貸歧視的案例，並歸納其發生的原因，再討論反歧視法中的受保護特徵，檢視人工智慧歧視中與受保護特徵使用的具體關係，最後再觀察我國相關法制規範。

第一節　人工智慧歧視

第一項　人工智慧與機器學習概述

人工智慧系統運用機器學習協助決策[17]，以替代人力的運算與審核。根據 OECD 對人工智慧的描述，可以依據其運作模式的不同分為兩種類型，一是有著明確規則的運作模式，由制定者事先針對可能產生的疑慮或欲防免的結果，訂出明確的規範與指令，此種模式通常較無爭議；另一類型則是機器學習[18]，機器學習的運作是人工智慧透過經驗學習並改進演算法以達到更佳的結果，1950 年代的人工智慧先驅 Arthur Samuel 對於機器學習的定義為：「無需明確編程即可賦予電腦學習能力的研究領域

[16] OECD (2020), OECD BUSINESS AND FINANCE OUTLOOK 2020: SUSTAINABLE AND RESILIENT FINANCE, OECD PUBLISHING, PARIS, https://doi.org/10.1787/eb61fd29-en.(last visited Jan 3, 2025).
[17] Sandra Wachter, *The Theory of Artificial Immutability: Protecting Algorithmic Groups under Anti-Discrimination Law*, 97 TUL. L. REV. 149, 1 (2022), https://www.tulanelawreview.org/pub/artificial-immutability (last visited Nov 20, 2023).
[18] Janneke Gerards & Frederik Zuiderveen Borgesius, *Protected Grounds and the System of Non-Discrimination Law in the Context of Algorithmic Decision-Making and Artificial Intelligence*, 20 COLO. TECH. L.J., 11(2020), https://ctlj.colorado.edu/wp-content/uploads/2022/06/PROTECTED-GROUNDS-AND-THE-SYSTEM-OF-NON-DISCRIMINATION-LAW-IN-THE-CONTEXT-OF-ALGORITHMIC-DECISION-MAKING-AND-ARTIFICIAL-INTELLIGENCE.pdf (last visited Nov 20, 2023).

doi:10.6342/NTU202500326

[19]。」而從銀行公會公布的《金融機構運用人工智慧技術作業規範》第一條及金管會公布的《金融業運用人工智慧指引》總則第一條，其對於人工智慧的定義為「透過大量資料學習，利用機器學習或相關建立模型之演算法，進行感知、預測、決策、規劃、推理、溝通等模仿人類學習、思考及反應模式之系統。」，相較前述 OECD 對人工智慧的定義，可以發現目前我國對於金融業中的人工智慧定義，係以必須包含利用機器學習或建立模型演算法為要件[20]。

機器學習又可分為監督式學習（supervised learning）、非監督式學習（unsupervised learning）、深度學習（deep learning）等子類別。監督式學習即為從標記資料中學習的機器學習，由人類標記的資料訓練演算法以產生正確結果[21]，監督式學習也是最普遍出現的[22]，其運用訓練資料使演算法知道著重的面向，在面對新輸入的資料時容易找到目標資料[23]； 非監督式學習則不同於監督式學習，會在沒有被事先標記的資料中尋找模型或趨勢，也因此非監督式學習可能可以找到人們預期之外的結果[24]；深度學習則像是具有許多層次的神經網路，其分層網路可以處理大量資料，並決定網路中每個環節的權重比例，深度學習是模仿人腦的運作方式為基礎，也因為層次和網路的複雜性，深度學習需要大量的運算能力[25]。

第二項　機器學習的不透明性

[19] Samuel, Arthur L., *Some Studies in Machine Learning Using the Game of Checkers*, IBM JOURNAL OF RESEARCH AND DEVELOPMENT 44:1.2, 210–229 (1959).
[20] 金管會（2024），前揭註 15，頁 3；銀行公會（2023），《金融機構運用人工智慧技術作業規範》，頁 1。
[21] Geeksforgeeks, *Supervised and Unsupervised Learning* (23 Sep, 2024), https://geeksforgeeks.org/supervised-unsupervised-learning/ (last visited Jan 4, 2025).
[22] Mit Tech. Rev., *Is AI Riding a One-Trick Pony?* (2017), https://www.technologyreview.com/s/608911/is-ai-riding-a-one-trick-pony/ (last visited Jan 3, 2025).
[23] Christopher K. Odinet, *Fintech Credit and the Financial Risk of AI*, U. IOWA LEGAL STUD. RESEARCH PAPER, No. 2021-39, 13 (2022).
[24] MIT Management, *Machine learning, explained* (Apr 21, 2021), https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained (last visited Jan 3, 2025).
[25] *Id.*

doi:10.6342/NTU202500326

機器學習產生的人工智慧的決策過程並不透明，論者常以「黑盒子」稱之[26]，即使是人工智慧的開發者也無法清楚了解其訓練資料是如何被人工智慧所運用，從而無法得知決策過程的細節。

Jenna Burrell 進一步提出了三種機器學習所存在的不透明性：第一，涉及國家或商業秘密所致的不透明性；第二，因技術盲點所致的不透明性；第三，因機器學習算法的特性及其應用所需的規模所致的不透明性[27]。整體而言，縱使人工智慧的程式碼被透明地揭露，但機器學習的演算過程及資料使用邏輯可能超出人類一般的認知，致使人類難以掌握資料運用間的關係，也難以實際檢查[28]，使得決策過程神秘且不直觀[29]。

從技術上討論機器學習的不透明性，學者 William Nicholson 及 Arti Kaur 將進一步其劃歸於三類原因：複雜性、不直觀性及保密性[30]：

一、 複雜性

機器學習模型的複雜性源自多種因素，包括非線性（nonlinearity）和不連續性（discontinuity）的特性，且相較於人工設計的模型而言有更多的變數（variables），每個變數對預測結果的貢獻難以估計，導致其模型運算系統極為複雜[31]。在某些情況下，特定的機器學習模型可以嘗試解釋潛在的系統複雜性，無論是在一般情況下或在特定情況下，然而這些解釋通常使用機器學習來詮釋機器學習，是典型的統計近似，而非完全精確的表達[32]。

---

[26] *See, e.g.*, Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORDHAM L. REV., 1090 (2018), https://ir.lawnet.fordham.edu/cgi/viewcontent.cgi?article=5569&context=flr (last visited Jan 3, 2025).
[27] Jenna Burrell, *How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms*, 3 BIG DATA SOC., 1 (2016), https://doi.org/10.1177/2053951715622512 (last visited Jan 3, 2025).
[28] Selbst & Barocas, *supra* note 26, at 1085,1089,1131.
[29] Selbst & Barocas, *supra* note 26, at 1089.
[30] William Nicholson Price II & Arti Kaur Rai, *Clearing Opacity through Machine Learning*, CO-AUTHORED BY ARTI K. RAI. IOWA L. REV. 106, NO. 2, 785 (2021), https://ilr.law.uiowa.edu/sites/ilr.law.uiowa.edu/files/2023-02/Nicholas%20Price_Rai.pdf (last visited Nov 20, 2023).
[31] Selbst & Barocas, *supra* note 26, at 1094-96.
[32] Price II& Rai, *supra* note 30, at 785.

doi:10.6342/NTU202500326

二、 非直觀性

非直觀性係指在機器學習的模型分析中，縱使透過直接分析找到資料輸入和結果的相關性，其相關性亦不易被人類解釋。William Nicholson 及 Arti Kaur 在文中提出了非直觀性的風險，其一為機器學習發現的相關性有虛假的可能，可能是來自於資料中並非真實的關係；其二是得以解釋輸入資料和結果間關聯性的變數可能是一個人類未知的潛在變數，導致人類無法解釋。

其提及了一則關於個人早餐的偏好實驗，實驗發現具有某種社會經濟地位的人，通常會選擇某種類型的早餐，同時也會規律地服藥。在此實驗中，事實上是因為早餐中的葡萄柚汁含有呋喃香豆素化學物質（furanocoumarin chemicals）抑制藥物代謝所需的酶（enzyme），從而影響了某些藥物的代謝。由此實驗發現，其之所以預測出特定人們的早餐喜愛偏好，事實上是因為有一個隱藏的潛在變數影響著預測結果，也就是因為這些特定人們有服藥的習慣，而早餐中常見的葡萄柚汁將影響藥物的代謝，所以這些人會避開含有葡萄柚汁的早餐類型，實際上影響因素是因「藥物代謝影響」，然而機器學習在預測時，未必能找到真正具有因果關係的真正原因，而可能找到其他具有相關性的變數，例如這些服藥的特定人們剛好都是具有某些社會經濟地位的人們，而產生了「某些社會經濟地位的人們傾向選擇某些特定類型早餐且會固定服藥」的預測結果，這樣的不直觀和潛藏在其中的實際因素，使輸入資料和結果間可能難以解釋，也加深了不透明性。

三、 保密性

人工智慧系統的不透明性，將不利於機器學習模型的訓練和進步，因此有提高透明度的需求。然而，演算法的相關技術資訊直接揭露將侵害開發人員的營業秘密及競爭上的優勢。保密性所造成的不透明即來自於人工智慧系統的提供者基於商業機密或競爭上優勢的考量，對其演算法的程式碼、相關參數、訓練資料、訓練過程、生成模

doi:10.6342/NTU202500326

型等技術上的資訊保密。由於關鍵的技術資料與模型的不透明，導致其他的機器學習模型開發與改進將有更高的成本[33]。

第三項 機器學習的公平性疑慮

機器學習的不透明性也引發了關於偏見與潛在歧視的問題[34]，例如在機器學習的資料探勘（data mining）中的資料問題，可能使用了帶有歧視性的資料，而導致演算法出現不樂見的歧視結果。演算法所使用的訓練資料亦可能有侷限性，特別是在使用歷史資料預測未來時，許多概念可能隨著時間的演進已有所變化[35]，例如過去對於性別和種族的偏見，可能因為歷史資料的使用，而將偏誤透過演算法影響未來的目標結果預測。如美國知名電商 Amazon 被指控其所開發的人工智慧招募工具即對女性有偏見[36]，該公司從 2014 年起開發電腦程式以審查求職履歷，為了實現以自動化的方式尋求人才，但該公司並未發現其開發的系統應用於篩選應聘資料時，並未保持性別中立，而原因是因為其訓練資料是來自於過去十年收到的應聘履歷，大多來自男性應聘者，而這也反映了科技產業大多以男性為主導的現象，因而導致使用該工具篩選時，男性應聘者成功的機率大於女性應聘者[37]。

人工智慧歧視的發生有許多原因，根據學者 Lepri Bruno 之見解，大致上可歸類為四種類型[38]：第一，輸入資料的使用偏差，常導致中性特徵之間接歧視發生，例如美國用於警務預測的系統容易導致犯罪率與郵遞區號的連結，進而導致居住於特定郵遞區號區域之居民在其他領域遭到不平等的對待，另一種對於輸入資料的使用偏差是以

---

[33] Price II& Rai, *supra* note 30, at 788.
[34] Robert Bartlett, Adair Morse, Richard Stanton & Nancy Wallace, *Consumer-Lending Discrimination in the FinTech Era*, 143 J. FIN. ECON., 30 (2022), https://doi.org/10.1016/j.jfineco.2021.05.047 (last visited Jan 11, 2025).
[35] Odinet, *supra* note 23, at 28.
[36] Reuters, *Insight - Amazon scraps secret AI recruiting tool that showed bias against women* (October 11, 2018), https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/ (last visited Jan 3, 2025).
[37] *Id.*
[38] Bruno Lepri et al., *Fair, Transparent, and Accountable Algorithmic Decision-Making Processes: The Premise, the Proposed Solutions, and the Open Challenges*, 31 PHILOS. TECHNOL. 611, 4 (2018).

doi:10.6342/NTU202500326

代理（proxy）方式針對受保護特徵歧視，如同前述之犯罪率與郵遞區號的連結關係，若該郵遞區號之地區通常居住著相同種族的人們，而實際上該系統是將犯罪率與特定種族相連結，導致了雖以「郵遞區號」為外觀，但卻實質指涉特定種族有較高的犯罪率的代理現象；第二，導因於偏差的資料輸入，當人工智慧將偏差資料所帶有的偏見在人工智慧中循環，會進一步加劇歧視的形成。這些偏見可能係基於人類固有的偏見，或該領域性質本來就帶有不公平的系統性不平等，如前揭 Amazon 的招募工具案例，即是因科技領域在過去的資料中，從業人員以男性居多，進而在以過去的應聘資料為訓練時，這樣的男女數量不平衡的狀態也成為了人工智慧模型被訓練的一部分，使得最後其預測的結果對女性產生偏見；第三，人工智慧本身的設計或結果即帶有歧視，對於特定群體產生偏見，此基本上為演算法的決策錯誤，不同於前述原因是來自於輸入資料的偏誤，人工智慧模型可能基於其設計者故意或無意，導致模型本身就不是公平的；第四，模型的誤用，人工智慧可能在不同的使用情境下被錯誤的使用，進而導致歧視的發生。

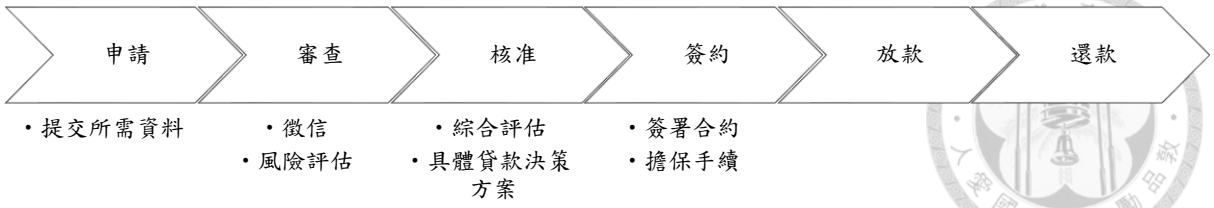
第二節　人工智慧信貸歧視

第一項　人工智慧信貸歧視的案例

信用貸款的流程主要分為申請、審查、核准、簽約等四階段，其中在審查與核准階段皆有涉及歧視風險的可能。在信用貸款的審查階段中，金融機構通常透過徵信系統查詢申請人的信用報告，以評估其信用狀況與還款能力，審查的資料可能包括歷史借貸紀錄、是否有逾期還款情形、信用評分等。除了徵信外，金融機構亦會針對申請人的薪資、職業穩定度、負債與否等資訊以判斷是否具有足夠的還款能力。經過綜合考量信用評估及風險後，決定是否核貸的階段即為授信階段[39]。

[39] 凱基銀行，〈信用貸款是什麼？信貸流程有哪些？貸款申請要多久？貸款流程申請必備資訊總整理！〉，（09/27/2024），https://www.kgibank.com.tw/zh-tw/kgibarticleshome/loan/pl-process（最後瀏覽日 01/01/2025）。

doi:10.6342/NTU202500326



【圖 1】信用貸款流程

資料來源：〈信用貸款是什麼？信貸流程有哪些？貸款申請要多久？貸款流程申請必備資訊總整理！〉，凱基銀行

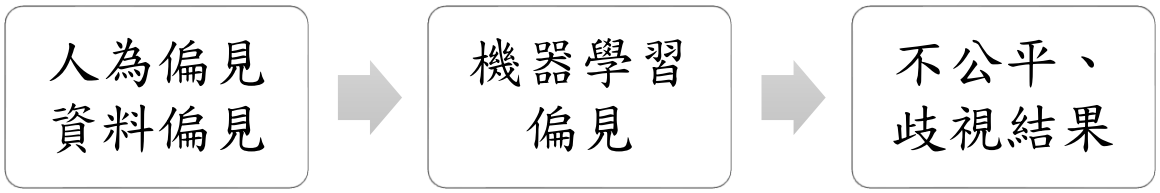
金融機構在信用貸款中運用人工智慧系統，主要係用以徵信業務中的信用評估及授信業務，藉由人工智慧系統的自動化、高效率及客觀準確性，得以節省人力成本並提高資料處理效率，用以評估潛在借款人的信用度、增強核貸決策流程並改善管理貸款組合，有助於金融機構精確、有效率地評估申請者之信用分數，降低誤判或人為歧視的風險[40]。近年各金融機構可見許多運用自動化核貸的信貸服務，我國例如合作金庫提出的「速速貸線上信用貸款」，透過線上申請並標榜符合要件者最快 30 分鐘即可入帳[41]；美國例如有人工智慧貸款平台 Upstart，貸款流程全程自動化[42]；中國例如廈門銀行專為臺灣人推出的「臺 e 貸」，透過標準化評分模型，以科學評估綜合申請者的個人資訊之自動化信用貸款[43]。人工智慧系統在信貸服務中的應用日益普及，自動化

[40] European Banking Federation, *EBF Position Paper on AI in the Banking Industry*, 8 (2019), https://www.ebf.eu/wp-content/uploads/2020/03/EBF-AI-paper-_final-.pdf. (last visited Jan 3, 2025).
[41] 黃有容，〈合庫銀將推「速速貸線上信貸」最快 30 分鐘入帳〉，（07/28/2023），《中時新聞網》，https://ynews.page.link/Jn8z5（最後瀏覽日：11/10/2024）。
[42] 美達衛，〈傳統銀行程序冗長 自動化借貸系統有創舉 AI 助攻信貸平台股價暴衝 5 倍〉，（09/29/2021），《財訊》，https://www.wealth.com.tw/articles/9e9060cd-9248-499e-801e-f1744f944c5d（最後瀏覽日：11/10/2024）。
[43] 徐迪，〈廈門銀行推出「臺 e 貸」 大陸首款臺胞專屬線上信用貸款產品〉，（09/06/2023），《台灣好報》，https://tw.news.yahoo.com/%E5%BB%88%E9%96%80%E9%8A%80%E8%A1%8C%E6%8E%A8%E5%87%BA-%E8%87%BAe%E8%B2%B8-%E5%A4%A7%E9%99%B8%E9%A6%96%E6%AC%BE%E8%87%BA%E8%83%9E%E5%B0%88%E5%B1%AC%E7%B7%9A%E4%B8%8A%E4%BF%A1%E7%94%A8%E8%B2%B8%E6%AC%BE%E7%94%A2%E5%93%81-083204181.html（最後瀏覽日：11/10/2024）。

doi:10.6342/NTU202500326

協助審核的流程不僅為金融機構降低成本、增加客觀上的準確度，也為申請的消費者帶來更快速、更便利的借貸服務。

人工智慧系統為信貸業務帶來效益，然而也具體化了人工智慧歧視。金融機構在信貸授信前，應先辦理徵信以調查申請人的信用狀況[44]，人工智慧系統於信用評分之應用將影響徵信、授信業務的審核與決策，人工智慧歧視問題也因此影響了徵信、授信的公平性問題，對申請人而言加劇了公平性風險。在人工智慧機器學習的運作中，需要大量的資料輸入，而資料本身是否夾雜著偏見？又或者資料是否基於過去人力審核而帶有歷史因素的偏見？是影響人工智慧信貸公平性的因素之一，因為將帶有偏見的資料輸入並經由機器學習後，將加深訓練資料帶來的偏差或錯誤，而產生潛在的風險，而有導致歧視的可能。人工智慧信貸歧視的另一個主要原因，來自人工智慧的不透明性。在徵信與授信的環節中，因人工智慧的複雜性、不透明性及金融機構的營業秘密保護，使得借款人和監管機關皆難以在過程中監管或介入，導致資料輸入與預測結果間的運作流程是難以理解的，若產生歧視結果時，也難以發現具體的問題癥結點。



【圖 2】演算法歧視的產生流程

資料來源：Algorithmic Discrimination, Sensitivenets[45]

[44] 如：中華民國銀行公會會員授信準則第 19 條第 1 項：「辦理授信案件，除法令另有規定外，於核貸前應先辦理徵信，未經辦理徵信者，不應核貸。」

[45] https://sensitivenets.com/what/

doi:10.6342/NTU202500326

人工智慧信用評分如同雙面刃，一方面提高信用評估的準確度，並擴大信用貸款的市場，使借貸者有更多機會申請。然而另一方面，人工智慧信用評分可能透過消費者的認知或行為弱點及有意的非法歧視意圖，導致新型態的不公平與低效率發生[46]。因此，在享受人工智慧系統帶來的優點時，應如何從監管或是立法規範為公平性風險提出解決方式，即為本研究的核心目標。

在信用貸款的申請審核中，借款人過去的信用紀錄是重要的風險判斷資訊之一，然而對於沒有信用紀錄的群體而言（例如學生、年輕人、新移民），信用紀錄的缺乏會導致無法準確評估其信用，形成信用隱形問題（problem of credit invisibility）[47]。具體而言，美國最常在信用評估流程中使用的模型，是由 Fair Isaac Corporation 創建的信用評分模型 FICO，透過 FICO 將借款人依據信用風險分成五個風險類別。但此類自動化評估程序對缺乏信用紀錄的群體十分不利，可能造成信用隱形或逐出效應。

針對信用記錄有限的客戶，信貸評估機構可能透過其他的個人資料評估信用風險，例如有金融科技公司轉而以其他資訊作為替代資料，即非傳統的「邊緣資料」（fringe data）[48]，例如在社群中或電信服務中的支付紀錄、健康紀錄、線上訂閱支付紀錄、消費者投訴紀錄、會員卡紀錄...等，作為其信用評估的考量[49]，如此可促進銀行服務信用紀錄不足人口的金融包容性[50]。

在人工智慧系統介入的信貸評估程序中，與個人相關的資料都可能成為信用評估的資料，透過廣泛資料以評估借款人，目的是做出更準確的信貸核保決策[51]。其根據個人條件與償還率區分，找出與逾期償還高相關性的特徵，列入信用評分的參考部

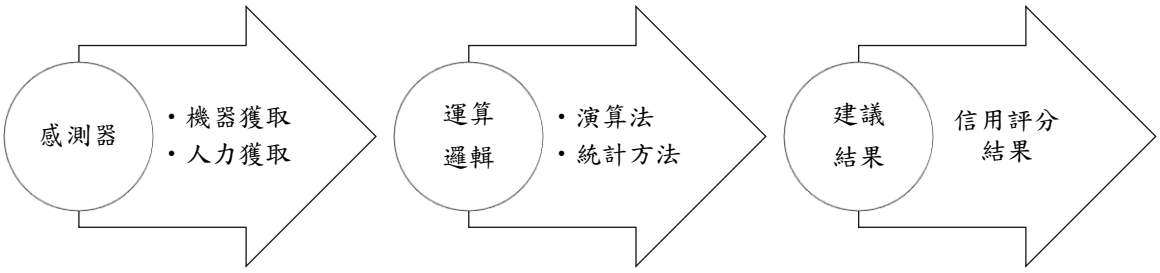
---

[46] Nikita Aggarwal, *Machine Learning, Big Data and the Regulation of Consumer Credit Markets: The Case of Algorithmic Credit Scoring*, AUTONOMOUS SYSTEMS AND THE LAW (2018), http://dx.doi.org/10.2139/ssrn.3309244. (last visited Jan 3,2025).
[47] Wu, *supra* note 4, at 135.
[48] *Id.* at 15.
[49] Ana Cristina Bicharra Garcia, Marcio Gomes Pinto Garcia & Roberto Rigobon, *Algorithmic Discrimination in the Credit Domain: What Do We Know about It?*, AI & SOCIETY, 13 (2023), https://doi.org/10.1007/s00146-023-01676-3. (last visited Jan 3, 2025).
[50] OECD, *supra* note 16.
[51] Wu, *supra* note 4, at 81.

doi:10.6342/NTU202500326

份。透過人工智慧系統的介入，排除了人力上的主觀因素影響，不僅更加客觀也提高了效率與準確性，亦得以避免歷史上基於個人品味（taste）而生的種族主義或性別歧視[52]。

根據 OECD 對人工智慧信用評分系統的解釋，應用於信用評分（credit scoring）的人工智慧系統運作模式為，首先根據感測器（sensors）所提供的資料，包括機器獲得的資料（借款人的個人資料及其歷史還款記錄）及人力獲得的資料（授信規則），透過運算邏輯（信用評分演算法、統計迴歸）後形成建議結果（信用評分），進而使執行者（金融機構）能得以出決策（是否核貸）。此系統的給定目標即為信用價值（credit-worthiness）[53]。



【圖 3】人工智慧中機器學習運作流程

資料來源：OECD

然而，人工智慧系統中的不透明性，可能增加信貸結果出現未預期的風險，導致具有偏見或歧視的不公平信貸。在美國消費者金融保護局（United States Consumer Financial Protection Bureau, CFPB）和美國司法部（Department of Justice, DOJ）發布

[52] Oren Bar-Gill, Cass R Sunstein, Inbal Talgam-Cohen, *Algorithmic Harm in Consumer Markets*, 15 JOURNAL OF LEGAL ANALYSIS, 31 (August 2023), https://doi.org/10.1093/jla/laad003 (last visited Jan 3, 2025)
[53] OECD (2019), SCOPING THE OECD AI PRINCIPLES, OECD DIGITAL ECONOMY PAPERS, https://www.oecd-ilibrary.org/docserver/d62f618a-en.pdf?expires=1589038170&id=id&accname=guest&checksum=4E099F3617BD00290FCC34E651F0920A(last visited May 29, 2024)；廖淑君（2012），〈人工智慧與普惠金融－淺析演算法於徵信／授信應用之金融消費者保護議題〉，《財金法學研究》，5 卷，1 期，頁 123。

doi:10.6342/NTU202500326

「關於打擊自動化系統中歧視和偏見的執法行動」聯合聲明中指出[54]：「當私人機構或公家機關使用自動化系統用於做出影響個人權利及權利的關鍵決策時，例如獲得工作、住房、信貸和其他商品之機會，這些系統可能有助於做出決策，然而也可能導致不合法的歧視結果。」

具體而言，看似客觀的人工智慧系統中，實則隱藏了新型態的公平性問題。首先，若用於訓練演算法的原始資料已帶有歷史上的偏見，那演算法即會帶有歧視[55]。如美國著名的紅線歧視（redlining），其係由社會學家約翰・麥克奈特（John McKnight）於 1960 年代提出的概念，源自聯邦政府和金融機構在地圖上用紅線劃出基於統計而拒絕核貸的社區，此類社區通常以非裔美人為主組成，導致金融機構寧願為低收入的白人提供貸款，也不願向中等收入甚至是高收入但居住於紅線社區的非裔美人提供[56]。在紅線歧視的案例中，統計外觀上其係以居住區域（郵遞區號）與呆帳率具有較高的相關性而劃定，但居住區域雖非法定受保護特徵，然而其背後所連結的實際上係針對種族進行的信貸歧視。

在人工智慧系統介入的徵信、授信流程中，亦有可能發生此類「數位紅線」（digital redlining）歧視的可能，美國聯準會亦曾表示此狀有違反《公平信貸法》的風險[57]。亦即儘管人工智慧系統固然不是有意透過中性特徵（居住區域）達到事實上歧視的結果，但在人工智慧系統追求較低的違約風險的目標下，其可能無意間將違約風

[54] United States Consumer Financial Protection Bureau & Department of Justice, Joint Statement On Enforcement Efforts Against, https://www.ftc.gov/system/files/ftc_gov/pdf/EEOC-CRT-FTC-CFPB-AI-Joint-Statement%28final%29.pdf (last visited Jan 1, 2025); FEDERAL TRADE COMMISSION, DISCRIMINATION AND BIAS IN AUTOMATED SYSTEMS, https://www.ftc.gov/system/files/ftc_gov/pdf/EEOC-CRT-FTC-CFPB-AI-Joint-Statement%28final%29.pdf (last visited Jan 3, 2025).

[55] *Id.* at 32.

[56] Investopedia, *What Is Redlining? Definition, Legality, and Effects*, https://www.investopedia.com/terms/r/redlining.asp (last visited May 29, 2024).

[57] Aba Banking Journal, *Fed's Barr highlights digital redlining risk resulting from AI tool*s, https://bankingjournal.aba.com/2023/07/feds-barr-highlights-digital-redlining-risk-resulting-from-ai-tools/. (last visited Nov 10, 2024).

doi:10.6342/NTU202500326

險與居住區域連結，進而導致居住區域的居民得到較低的信用評分或是核貸機會，產生實質的歧視風險。

以下列舉若干人工智慧信貸歧視的案例。

第一款　Wells Fargo 房貸信用評分歧視案

2022 年彭博社（Bloomberg）的一則調查發現，Wells Fargo 在抵押貸款融資中，對於有色人種申請者的核准比率低於其他種族的申請者[58]，導致原告 Aaron Braxton 等人向 Wells Fargo 提起集體訴訟[59]。原告在訴訟主張，Wells Fargo 使用內部信用評估系統（Enhanced Credit Score）用以預測申請人違約的可能性並評分，但該系統的演算法將非裔和拉丁裔申請人列入風險較高的等級，使其須經過更多審核，並導致更高的不利貸款結果[60]。

第二款　海軍聯邦信用社歧視黑人貸款案

2023 年 12 月 17 日，Laquita Oliver 和 Cerelle Jacob 為首代表向美國海軍聯邦信用社（Navy Federal Credit Union）提起集體訴訟，指控其因種族因素拒絕房屋貸款，違反公平住居法與平等信貸機會法。在訴訟主張中，原告引用了 CNN 的一則分析報告[61]，該分析報告是基於《住房抵押貸款揭露法案》（The Home Mortgage Disclosure Act, HMDA）所公開的資訊進行分析，發現在 2022 年房屋抵押貸款的申請中，海軍聯邦信

[58] Bloomberg, *Wells Fargo Rejected Half Its Black Applicants in Mortgage Refinancing Boom* (Mar. 11, 2022), https://www.bloomberg.com/graphics/2022-wells-fargo-black-home-loan-refinancing/?utm_source=website&utm_medium=share&utm_campaign=copy (last visited Nov 10, 2024).
[59] *Braxton v. Wells Fargo Bank, N.A. a Delaware corporation, 3:22-cv-01748, (N.D. Cal.)*.
[60] Housing Wire, *Mortgage Underwriting Algorithm at Heart of Wells Fargo's Racial Disparity Lawsuit* (June 14, 2024), https://finance.yahoo.com/news/mortgage-underwriting-algorithm-heart-wells-161043355.html. (last visited Nov 10, 2024).
[61] CNN, *The Nation's Largest Credit Union Rejected More Than Half Its Black Conventional Mortgage Applicants* (Dec. 14, 2023), https://www.cnn.com/2023/12/14/business/navy-federal-credit-union-black-applicants-invs (last visited Nov 10, 2024).

doi:10.6342/NTU202500326

用社核准了 77.1%的白人申請人、55.8%的拉丁裔申請人和 48.5%的非裔申請人，其中非裔和白人申請人之間的差距高達 28.6 個百分點，是金融機構中差距最大的[62]。

根據調查顯示，海軍聯邦信用社的抵押貸款過程係由自動決策系統審核，而海軍聯邦信用社對於其系統中的演算法運作始終保密，不願揭露與審核相關的資訊，足見該演算法與歧視的貸款審核結果間有著高度關聯性。該系統使用的各種統計特徵，例如教育、郵遞區號等，實質上可能被作為「種族」的代理特徵（Proxy），故儘管原告 Laquita Oliver 和 Cerelle Jacob 皆擁有良好的收入、職業和信用分數，卻仍然因為「種族」因素得到不利的貸款結果[63]。

第三款　Apple Card 案

美國高盛銀行（Goldman-Sachs）與蘋果公司推出的 Apple Card，也是使用演算法決策而產生歧視疑慮的案例之一。2019 年 11 月，高盛與蘋果所推出的 Apple Card 客戶表示，其信貸評估演算法在分配信貸額度時出現性別歧視，男性獲得的信貸額度明顯高於相同條件的女性，高盛針對該事件發表聲明，表示其未曾且未來也不會以性別、種族或任何其他法律禁止的受保護特徵做出信貸價值的判斷[64]。紐約州金融服務部經過調查後亦表示，Apple Card 企業並沒有違反公平信貸法（ECOA），因為其在開發演算法時並沒有使用性別資料[65]。

有學者為了找出 Apple Card 沒有使用性別資料卻仍然導致性別歧視的原因，透過金融科技公司 Home Credit 所提供的公開、非抵押貸款的資料庫建立機器學習模型，試

---

[62] *Id.*
[63] BET, *Black Home Mortgage Borrowers Sue Navy Federal Credit Union For Discrimination* (2023), https://www.bet.com/article/vlo62w/navy-federal-credit-union-lawsuit-racial-discrimination. (last visited Nov 10, 2024).
[64] AI Incident Database, *Incident 92: Apple Card's Credit Assessment Algorithm Allegedly Discriminated against Women* (2019), https://incidentdatabase.ai/cite/92/ (last visited Nov 10, 2024).
[65] New York State Department of Financial Services, *Report on Apple Card Investigation* (2021), https://www.dfs.ny.gov/system/files/documents/2021/03/rpt_202103_apple_card_investigation.pdf (last visited Jan 3, 2025).

doi:10.6342/NTU202500326

圖找出統計和機器學習歧視的驅動因素[66]，進而發現 Apple Card 案例的不當結果可能是來自歷史資料的偏差。從模擬資料中發現，在其他條件不變之下，女性比男性更適合作為借款人，工作經驗豐富的人比工作經驗較少的人更適合作為借款人，但資料顯示女性的平均工作經驗通常少於男性。此外，訓練資料庫中的歷史借款人的資訊中，在全球範圍內 80%為男性及 20%為女性。由於女性借款人在過去的借款人中佔少數，故當演算法無法使用性別資料時，工作經驗資深的借款人較容易獲得信用，導致演算法預測的「平均」借款人行為將更傾向於男性，加劇了對女性的歧視結果[67]。

然而，儘管在使用模擬資料的研究結果中，推測出導致 Apple Card 演算法模型導致性別歧視的原因來自於歷史借款人中的工作經驗及性別比例問題，但在前述 Apple Card 的性別歧視案例中，男性和女性都具有相同良好的個人條件，同樣都具有工作經驗，就此觀之，該模擬研究以女性的平均工作經驗少於男性進而導致對女性借款人有歧視結果的論述，仍無法就此案例有合理的解釋。

第二項　人工智慧信貸歧視的原因分類

金融機構作為貸方本應中立，對於任何借貸者不應在核貸程序對其有主觀偏見。實務上，借貸者所受到的信用歧視可能有兩種類型[68]，一是基於金融機構的偏好的主觀歧視（taste-based discrimination），即金融機構基於個人偏好而非客觀利潤最大化的方式進行核貸決定，導致不公平對待的情形；另一種類型則為統計上的歧視（statistical discrimination），即信貸機構基於客觀風險評估而產生的歧視[69]。理想的狀態下，金融機構在決定是否核貸給借款人以及決定貸款的利率時，應基於其個人的信用風險情形而做出合理的評估，然而若其並非基於個人的信用風險，而是基於偏好對

---

[66] Knowledge, *How Anti-Discriminatory Measures Can Worsen AI Bias* (2023), https://knowledge.insead.edu/operations/how-anti-discriminatory-measures-can-worsen-ai-bias. (last visited Nov 10, 2024).
[67] *Id.*
[68] Katja Langenbucher, *supra* note 1010, at 10.
[69] *Id.* at 10.

doi:10.6342/NTU202500326

某些群體有不平等的對待時，此時其所受到的貸款結果並非是來自於信用、違約風險的考量，而是對於某些群體的歧視，除了違反了反歧視的普世原則外，對於金融機構而言也沒有達到客觀上追求利潤最大化，所以相較之下，前述基於恣意的主觀偏好歧視是確保信貸公平性所需防免的[70]。

如前揭學者 Lepri Bruno 對人工智慧歧視之分類，將其分為輸入資料的使用偏差、偏差的資料輸入、本身即帶有歧視的人工智慧模型及人工智慧模型的誤用等原因[71]，以下從資料輸入因素的面向及模型因素面向，分別探討在人工智慧信貸中具體的歧視型態。

第一款　輸入資料的使用偏差與系統性歧視

人工智慧信貸之所以產生不公正的信用評估結果，主要原因之一是來自於輸入資料的使用偏差，如前述，信用歧視可以被分為金融機構基於偏好的主觀歧視，或是為了追求客觀利益最大化的統計歧視。在傳統信貸服務中，已規範金融機構不得針對法規明訂的受保護對象為歧視行為，同樣地在運用人工智慧模型的信貸服務中，當然禁止其係基於主觀偏好而對法規明定的受保護對象為歧視。然而，在金融機構為了達到利益最大化且儘可能降低風險的目標時，縱使未直接針對受保護對象有基於主觀偏好的歧視行為，然而可能因為這些受保護對象的個人條件與還款風險有所連結，而導致人工智慧系統在追求效率最大化的目標時，可能使用了某些相對中性的資料，但實際上仍然是基於對於受保護對象所為的歧視[72]。

另一種輸入資料導致歧視原因是輸入資料的不足，導致金融機構無法對借款人有正確的評估而演變成系統性歧視，亦可能是輸入資料本身帶有偏見而不斷訓練人工智慧模型，加強的系統性的歧視。理論上，人工智慧系統應可避免人為主觀的偏見歧

[70] *Id.* at 10.
[71] Bruno Lepri, *supra* note 38, at 4.
[72] Cornell University, *The Fairness of Credit Scoring Models*, 6, 12 (2024), http://arxiv.org/abs/2205.10200. (last visited Feb 20, 2024).

doi:10.6342/NTU202500326

視，然而在運用的實務上，人工智慧僅能從借貸者過去的還款或其他信用表現評估[73]，故缺乏足夠的歷史資料的借款人，將可能產生不準確或有異狀的預測結果。除了個人的歷史資料外，各種歷史上的偏見資料也可能影響機器學習的判斷，例如某些被社會價值觀認定有高風險違約的特徵，可能在機器學習中影響核貸決策，導致系統性的歧視發生。

訓練資料的來源可能是被動、主動或是自動生成的，資料來源時常也可能來自於使用者的個人生活足跡。在學者阿格瓦等人（Agarwal et al.）及貝爾格等人（Berg et al.）的文章中發現，手機是重要的資訊來源，從手機選擇的使用品牌（如 Apple 和其他Android系統手機）、手機使用的資料數據（如安裝的應用程式、社交軟體中的通訊數量、頻域等）結合個人基本條件，進而預測在信貸中的違約風險[74]。然而在此來源的資料中，可能因年齡、貧富的資源落差產生資料不具有公正性而有資訊落差的疑慮，進而使得訓練結果有所不公平。

第二款　人工智慧模型的缺陷與誤用

在信貸服務中，是否核貸的關鍵因素在於借款人的信用分數[75]，其核心即為借款人的信用風險。信用分數的計算是透過統計模型的運算，此模型對於各家金融業者而言為商業機密[76]，目前常用以信用評估的機器學習演算法分別是 Random Forest[77]、

[73] *Id.* at 12.
[74] Katja Langenbucher, *supra* note 10, at 12.
[75] U.S. PUBLIC INTEREST RESEARCH GROUP (USPIRG) AND THE CENTER FOR DIGITAL DEMOCRACY (CDD), EXPANDING ACCESS TO CREDIT THROUGH ONLINE MARKETPLACE LENDING, COMMENTS TO THE U.S. DEPARTMENT OF THE TREASURY (2015), https://www.democraticmedia.org/sites/default/files/field/public/2015/uspirg_cdd_marketplacelendingrf_fnal30sept2015.pdf (last visited February 13, 2024).
[76] Wu, *supra* note 4, at 84.
[77] Nadège Grennepois, Anca Maria Alvirescu, Margaux Bombail, *Point of View: Using Random Forest for Credit Risk Models*, DELOITTE RISK ADVISORY (Aug. 2019), https://www2.deloitte.com/content/dam/Deloitte/sg/Documents/financial-services/sg-fsi-machinelearning-credit-risk.pdf (last visited Jan 5, 2025).

doi:10.6342/NTU202500326

Artificial Neural Networks[78]和 Boosting[79]，如常見機器學習演算法，其步驟首先為收集並清理資料，再將資料分為訓練資料集及測試資料集，進而根據機器學習演算法使用訓練資料集訓練預測模型，最後使用測試資料集驗證模型。若人工智慧模型本身有缺陷而不公正，或被金融機構誤用而導致模型產生不公平的訓練結果，則可能造成另一種人工智慧信貸歧視的原因。

信貸歧視的結果，可能是拒絕核貸、要求更高的貸款利率、給予更低的信用額度或需要更多的擔保等不利後果[80]。在人工智慧信貸歧視中，除非是人工智慧系統直接以受保護特徵對借貸人有不公平的對待而得受公平信貸法規範，若是其他如上述輸入資料的代理使用或系統性歧視，通常難以被現有的法律規範所規範。

第三項　反歧視與公平性要求

運用人工智慧系統進行預測或決策時，具有高效率、客觀化的優點，但機器學習的決策也有可能出現歧視或不公平的偏見。從經濟學的觀點而言，歧視可分為基於品味的主觀歧視，及基於統計結果的客觀歧視[81]。主觀歧視的「品味」來自於個人的偏見或偏好，學者認為，若此品味的維持需要付出額外的代價時，只要超過其得以忍受的臨界值時，市場的力量即會消除此主觀歧視[82]；而統計歧視係因對個人缺乏必要資訊以計算風險，故而將個人所屬的群體風險歸因於個人，從而產生歧視[83]，例如某群體的歷史信用紀錄不佳，會使屬於該群體的個人在統計上預測未來的違約率較高[84]。

---

[78] Oracle Ai &Data Science Blog, *Using the Artificial Neural Network for Credit Risk Management* (Jan. 23, 2019), https://blogs.oracle.com/ai-and-datascience/post/using-the-artificial-neural-network-for-credit-risk-management (last visited Jan 5, 2025).
[79] Medium, *A Quick Guide to Boosting in ML* (Mar. 21, 2018),https://medium.com/greyatom/a-quick-guide-to-boosting-in-ml-acf7c1585cb5/ (last visited Jan 5, 2025); Wu, *supra* note 4, at 79.
[80] Cornell University, *supra* note 72, at 4.
[81] Galina Andreeva, Jake Ansell & Jonathan Crook, *Impact of Anti-Discrimination Laws on Credit Scoring*, 9 J. FINANC. SERV. MARK. 22, 25 (2004).
[82] Becker, G. S. (1971) 'The Economics of Discrimination', University of Chicago Press.
[83] Talia B. Gillis, *The Input Fallacy*, 106 MINN. L. REV. 1175, 18 (2022). https://scholarship.law.columbia.edu/faculty_scholarship/3312 (last visited Jun 12, 2024); Katja Langenbucher, *supra* note 10, at 17.
[84] Andreeva, Ansell, and Crook, *supra* note 81, at 25.

doi:10.6342/NTU202500326

人工智慧可能產生的歧視樣態相當多元，除了針對傳統反歧視法上規範受保護特徵的歧視分類之外（如以種族、性別、年齡...等分類），亦有可能出現針對非受保護特徵的歧視，也就是任何特徵都可能成為歧視的相關因素，此又稱為「中性特徵」之歧視，而若此中性特徵的使用，實際上會連結至受保護特徵分類對象時，其所造成的歧視結果應該被禁止，因其仍然構成對受保護特徵分類對象的不公平，然而因其中性的特徵外觀，故其所造成的歧視相較直接以受保護特徵歧視的行為難以被察覺，且因其表面上可能為各式各樣的態樣，亦難以如受保護特徵一樣被法律明定為保護分類，而無法被現行的反歧視法制保護囊括。

以下自歧視的對象分類，就法定受保護特徵分類之歧視及非法定受保護特徵分類之歧視為討論。

第一款　受保護特徵的歧視

針對受保護特徵的歧視，基於主觀上是否具有故意，可以分為直接歧視（direct discrimination）和間接歧視（indirect discrimination）兩個種類。所謂直接歧視係為法律規範的受保護分類標準所為之差別待遇，如司法院大法官解釋第 728 號解釋文多數意見，認為若未以生理性別作為受保護分類標準，則不構成直接歧視；而間接歧視即指的表面上對法律規範受保護分類標準的對象沒有歧視，但實際上仍對其產生歧視的效果（discriminatory effect），如司法院大法官解釋第 760 號解釋文提到的系統性之不利差別待遇[85]，因為縱使表面上的行為可能是中性措施，但可能忽略了結構或歷史上既存的不平等，而加深不平等導致間接歧視。

歧視的發生原因，可能來自於訓練資料受到歷史的決策影響，常見的是基於性別和種族，導致機器學習後的預測結果也帶有偏見。另一可能是演算法以自動化的方式發現訓練資料中的相關性，進而重現此偏見[86]。如 Sweeney 討論關於 Google 廣告中的

[85] 司法院釋字第 760 號，理由書第 10 段。
[86] *Id.* at 10.

doi:10.6342/NTU202500326

歧視現象[87]時，發現在 Google 搜尋引擎中輸入非裔美國人的名字，搜尋結果會顯示暗示有逮捕紀錄的廣告，而若係輸入白人名字，暗示逮捕紀錄廣告的頻率較低，此似乎是出於 Google 演算法於分析哪些廣告的點擊率較高時承襲了人群們偏好背後隱藏的種族歧視[88]。

各國既有的反歧視法規已明列許多明定禁止歧視的受保護特徵，例如美國《公平住房法》（FHA）禁止以「種族、膚色、宗教、性別、身心障礙、婚姻狀況或身心障礙」拒絕貸款；美國《平等信用機會法》（ECOA）禁止任何貸款人在信貸交易的任何流程或方面，對申請人有以下歧視，若有則屬違法：「1. 基於種族、膚色、宗教、國籍、性別或婚姻狀況或年齡；2. 因為申請人的全部或部分收入來自任何公共援助計劃；或 3. 因為申請人曾善意行使本章規定的任何權利[89]」。行為人如涉及故意利用已被明列的受保護特徵之歧視，會構成「直接歧視」或「差別對待歧視」，若非故意則為「間接歧視」或「差別影響歧視」[90]。

受保護特徵的選定理由，通常係基於保障個人不因先天或後天無法改變之特徵而受有不公平待遇，學者 Sandra 進一步歸納了四種受保護特徵的選定理由[91]：

一、不變性與選擇

不變性指的是不得因他人無法控制或改變的特徵，而使其處於不利的地位而受有不公平的對待，而這些特徵係與道德無關的，傳統上認為無法改變的特徵包括性別、種族、身心障礙等，這些也是反歧視法中常見的受保護特徵[92]。

二、相關性、任意性與優點

[87] Latanya Sweeney, *Discrimination in Online Ad Delivery*, HARVARD UNIVERSITY, 1 (2013).
[88] *Id.* at 3.
[89] 15 U.S.C. §1691.
[90] Wu, *supra* note 4, at 94.
[91] Wachter, *supra* note 17, at 16–28.
[92] *Id.* at 17.

doi:10.6342/NTU202500326

若歧視行為所基於的理由係來自於與其行為無關的受保護特徵時，例如以種族決定是否核貸或給予不利的貸款條件，這樣的受保護特徵使用被稱為是任意決定的，也就是使用受保護特徵與歧視行為間是無關的[93]。但實際上，行為人在使用看似無關的特徵為歧視行為時，亦有可能是具有相關性的，舉例而言，金融機構以種族作為判斷授信的依據時，可能是基於特定種族統計上可能帶有較高的風險比例，因此以此為做出不利對待的動機[94]，從這樣的角度而言，其對於受保護特徵的使用仍然是具有相關性，甚至可以說是理性決定的結果。因此，基於所為的歧視型為，也因納入制定受保護特徵分類時的參酌因素。

三、歷史壓迫或結構性劣勢

有學者認為，若干反歧視法保護的特徵亦可能不具有共通的邏輯，之所以列為受保護特徵，其目的是保護具有共同歷史經歷的群體，避免其繼續受到歷史上不公平的延續迫害[95]。從現行反歧視法中常見的受保護特徵觀察，大多是為了避免延續歷史上的不平等，例如種族和性別的歧視。然而在考慮面對新型態的歧視時，其可能未必有共同的歷史壓迫經驗，且受到壓迫和不平等待遇除了歷史原因外，更多時候可能來自於文化上的因素，例如歧視肥胖者、歧視特殊疾病（如愛滋病）患者等[96]，因此，認定是否為受保護群體時，更應著重的是該特徵群體是否有受到社會上、群體中持續的壓迫。

四、社會顯著性

另一個作為受保護特徵分類的依據是群體的社會顯著性，所謂的社會顯著性指的是受保護群體在社會上深刻地受到不同的對待和認知[97]，且具有該受保護特徵的

[93] *Id.* at 19.
[94] *Id.* at 19.
[95] Deborah Hellman & Sophia Moreau, eds., *Philosophical Foundations of Discrimination Law*, 169-170 (2013), SUFFOLK UNIVERSITY LAW SCHOOL RESEARCH PAPER NO. 13-27, https://ssrn.com/abstract=2303903 (last visited Feb 17, 2024).
[96] Wachter, *supra* note 17, at 23.
[97] David Wasserman, *Discrimination, concept of*, 1 ENCYCLOPEDIA OF APPLIED ETHICS, 807 (1998).

doi:10.6342/NTU202500326

群體自身也對自己具有這樣的特徵有所意識[98]，受保護群體間可能有「團結、連貫、認同、共同的歷史、語言或文化」的特性，可能是社會上有明顯不利地位的群體。然而在人工智慧歧視中，被歧視的群體可能是基於新型態的歧視特徵，所以未必具有如此受保護群體的顯著性或認知，若這樣的群體不利地位是分類受保護特徵常見的參酌因素時，那麼人工智慧歧視的新型態歧視特徵將難以以此分類[99]。

第二款　非受保護特徵的歧視

非受保護特徵的歧視即使用「中性特徵」的「代理」方式為歧視行為，行為人可能是故意也可能是無意為之，例如，演算法發現居住在某些特定區域人民的郵遞區號和種族具有相關性時，若某公司想要歧視一定種族的求職者時，可迂迴透過限制郵遞區號以達成歧視種族的結果[100]。於此種情形，中性特徵實際上代替受保護特徵被使用，進而實質上仍構成不公平的歧視結果。

在非受保護特徵中，又可區分為兩類，一類是以人類可理解特徵的代理歧視，如上述的郵遞區號與種族的關聯性；另一類則是人類難以理解的特徵，來自演算法分類的特徵，如以網站上滑鼠的移動軌跡的反應時間[101]，觀察是否具有隱性歧視的存在，即係所謂「內隱刻板印象」（implicit stereotype）的概念[102]，也就是在語義記憶的聯想網路系統中，社會概念（例如種族或性別）與屬性（例如正面或負面）是緊密結合在一起的，根據這個理論，儲存在語義記憶中的刻板聯想（例如「黑人」和「負面」）可能會被自動啟動而引發隱性刻板印象效應，也就是測試中發現，人們看到不同種族的圖片時，其滑鼠點擊移動軌跡的時間有所不同，而顯現了隱性的歧視[103]。

---

[98] Kasper Lippert-Rasmussen, *Born Free And Equal?: A Philosophical Inquiry Into The Nature Of Discrimination*, 34, OXFORD UNIVERSITY PRESS (2014).
[99] Wachter, *supra* note 17, at 28.
[100] Gerards and Zuiderveen Borgesius, *supra* note 18, at 11.
[101] Wachter, *supra* note 17, at 6.
[102] Di Palma M, Carioti D, Arcangeli E, et al, *The biased hand. Mouse-tracking metrics to examine the conflict processing in a race-implicit association test*, 17(7) PLOS ONE, 2 (2022), https://pmc.ncbi.nlm.nih.gov/articles/PMC9328548/ (last visited Jan 3, 2025).
[103] *Id.*

doi:10.6342/NTU202500326

第三款　受保護特徵與中性特徵歧視在人工智慧歧視中的態樣

從保護特徵的使用情境不同，人工智慧信貸歧視有不同的出現態樣，學者 Katja 將造成歧視結果的情景分成三種類型討論[104]，提出人工智慧信貸歧視與受保護特徵運用間的關係。

類型一為混合性動機歧視（mixed motive discrimination），此係指人工智慧直接使用了受保護特徵和其他的中性變數，且使用者聲稱受保護特徵是測量信用違約風險的良好指標，故而使用之。例如在信貸決策中，同時考慮了性別（受保護特徵）和其他非歧視的中性因素，在反歧視法架構下仍然會因為使用到受保護特徵而違法[105]。

類型二指的是，貸款人並未在人工智慧模型中使用任何受保護特徵，因此無法從其人工智慧模型中找到任何違反反歧視法的行為，然而其雖然沒有直接使用受保護特徵，但可能刻意選擇與這些受保護特徵高度相關的中性特徵，也因此這仍然導致對此些受保護群體有不利影響[106]。例如金融機構在進行信貸決策時，可能想針對某些特定種族為不利的決策，但因種族是受保護特徵而不得為之，其可能改以該種族常聚集居住的地區郵遞區號，作為決策的考量分類，實際上其仍然是刻意對該種族為不公平的對待。

類型三則是行為人使用人工智慧模型運算時，雖然並未故意選擇相關變數，但可能預見人工智慧模型可能會使用這些變數，並且這些變數可能會導致歧視性結果。儘管行為人沒有直接選擇受保護特徵，但這樣消極的行為仍然可能被視為間接歧視，因為其對模型的結果仍有一定的預見性[107]。例如如同前述類型二的例子，金融機構可能使用了一些看似中立的變數（如地理位置或薪資收入），但這些變數可能與受保護特徵（如種族或性別）存在隱含的相關性，從而導致某些群體在核貸過程中受到不利影

[104] Katja Langenbucher, *supra* note 10, at 21.
[105] *Id.* at 22.
[106] *Id.* at 29.
[107] *Id.* at 29.

doi:10.6342/NTU202500326

響，不同於類型二是金融機構刻意使用與受保護特徵相關聯的中性特徵，此類型的行為人係明知演算法可能產生此結果，卻不積極避免[108]。

第三節　我國法下的公平信貸規範與人工智慧信貸歧視

本節試圖梳理我國現行法制，是否有公平信貸或金融反歧視之相關規範，得以作為金融服務中的公平性要求之依據。另外，本節亦從金融機構應用人工智慧模型時，是否有應遵守的規範要求，而監管單位對於使用人工智慧的定義及規範方式又是如何？以下分別討論之。

第一項　公平信貸與金融反歧視規範

我國現行金融法規對於信貸歧視並無獨立的規範。例如《銀行法》第 34 條之 1：「銀行辦理授信，應訂定合理之定價，考量市場利率、本身資金成本、營運成本、預期風險損失及客戶整體貢獻度等因素，不得以不合理之定價招攬或從事授信業務。」但其僅係針對定價的合理性規範，並非反歧視規範。

又中華民國銀行商業同業公會全國聯合會（下稱「中華民國銀行公會」）訂立之《中華民國銀行公會會員授信準則》第 6 條有規定：「各級授信人員與客戶洽談應保持懇切之態度，對受理申貸案件所應徵提之資料應充分告知客戶，必要時得協助其依照規定格式填具申請書，並應秉持公正客觀之立場審查」、第 20 條第 1 項：「辦理授信業務應本安全性、流動性、公益性、收益性及成長性等五項基本原則，並依借款戶、資金用途、償還來源、債權保障及授信展望等五項審核原則核貸之」，惟其僅提出審核信貸之原則方針，並泛泛要求公正客觀審查，但仍未具體規定歧視禁止。

《金融消費者保護法》中亦無明文反歧視規範。其第 7 條第 1 項中之「金融服務業與金融消費者訂立提供金融商品或服務之契約，應公平合理、平等互惠及誠信原

[108] Fuster, Andreas and Goldsmith-Pinkham, Paul S. and Ramadorai, Tarun and Walther, Ansgar, *Predictably Unequal? The Effects of Machine Learning on Credit Markets*, 77 JOURNAL OF FINANCE, 8 (2022), http://dx.doi.org/10.2139/ssrn.3072038 (last visited Nov 5, 2024).

doi:10.6342/NTU202500326

則」之規定，該條規範要求金融業者和消費者所訂立的契約應符合公平合理原則，然而歧視問題除了契約內容的公平性外，對於金融業者是否決定與消費者締約，或是不同的締約對象間是否受到公平的對待，都可能產生歧視的風險，也因此，該條規範似乎得作為要求金融機構於提供金融服務時應注意公平性原則規範依據，但實際上並不足以囊括、防免金融服務中各種歧視的發生樣態。

目前我國法有明文提及「禁止歧視」規範之信貸相關法規，勉強僅見於《銀行業金融友善服務準則》第 2 條：「為提供身心障礙人士金融友善服務，應依不同類別身心障礙人士之需求，提供適當之友善服務措施，其範圍應包括環境、溝通、服務、商品、資訊、 檢核與統計及公告等無障礙措施，不得有歧視性之行為」、第 4 條第 1 項：「金融機構對身心障礙人士臨櫃提供金融商品或服務，應充分告知所需之資料，協助其填具相關申請書據並提供完整商品資訊，如依法有見證等需求，應尊重身心障礙者之選擇，給予協助，並應秉持公正客觀與不歧視之立場進行審查。」

第二項　金融業應用人工智慧規範

第一款　金融機構運用人工智慧技術作業規範

《金融機構運用人工智慧技術作業規範》是由中華民國銀行商業同業公會全國聯合會訂定，制訂的目的是為了強化金融機構運用人工智慧技術辦理銀行業務的客戶資料保護及銀行風險控管[109]，其規範對象的效力僅及於銀行公會的會員，定位屬於自律規範[110]。在此規範中所指的人工智慧，係透過大量資料學習之機器學習、演算法模型，而金融機構將人工智慧運用在提供消費者金融商品建議、提供服務且影響消費者權益或對營運有重大影響時，須符合此作業規範的要求[111]；而《金融業運用人工智慧（AI）

[109]《金融機構運用人工智慧技術作業規範》第一條

[110] 李佳熹，〈金管會 AI 指引、銀行公會 AI 規範發布，三大重點一次看懂！〉，《財團法人資訊工業策進會數位轉型研究院》，https://www.find.org.tw/index/indus_trend/browse/875b2d221ea9f38c4fc418f3d0089ead/（最後瀏覽日：05/08/2024）。

[111]《金融機構運用人工智慧技術作業規範》第二條

doi:10.6342/NTU202500326

指引》則是由金融監督管理委員會（下稱金管會）公布，其規範對象不限於銀行，擴及於所有金融機構，其指引中指出其性質為行政指導，亦即此指引並無強制力，僅能作為金融機構運用人工智慧系統時的行為參考。

金管會於 113 年 3 月 14 日同意備查中華民國銀行商業同業公會全國聯合會訂定之《金融機構運用人工智慧技術作業規範》[112]，全文共 16 條。第一條將人工智慧定義為「透過大量資料學習，利用機器學習或相關建立模型之演算法，進行感知、預測、決策、規劃、推理、溝通等模仿人類學習、思考及反應模式之技術」。

關於公平性，規範第 6 條第 1 項要求金融機構於運用人工智慧時，在演算法設計、開發、資料蒐集、訓練資料選擇、處理、模型建置/生成/優化，及後續應用於金融服務過程中，應採取措施以符合金融服務業公平待客原則；第 2 項則列出了面對公平性問題時，應考量的受保護特徵：「以下資料參數得評估是否納入演算法判斷，如：姓名、法律無限制或禁止之年齡、所有生理特徵（包含且不限於身高、體重、性別、膚色、髮量、肢體障礙等）、所有非涉及心神喪失致無法自主理解該金融商品判斷能力之疾病，並應就資安、法遵及風控等層面評估風險，依內部程序辦理。」

除明文規定公平性的要求外，規範第 7 條第 2 項中要求解釋人工智慧技術並提高模型的可解釋性；第 10 條要求資料文件的保存義務，涉及第三方機構時亦需留存查驗供查驗，並需明定問責架構；第 13 條說明了此規範將人工智慧系統以「風險」為劃分，如同歐盟《人工智慧法案》依據營業規模及運用人工智慧技術之程度建立適當之風險管理及定期檢視機制，亦屬維持人工智慧公平性相關之配套措施。

第二款　金融業應用人工智慧指引草案

[112] 金管銀國字第 1120237675 號函

doi:10.6342/NTU202500326

我國金管會於 112 年 10 月 17 日發布《金融業運用人工智慧（AI）之核心原則》與相關推動政策[113]，提出「建立治理及問責機制」、「重視公平性及以人為本的價值觀」、「保護隱私及客戶權益」、「確保系統穩健性與安全性」、「落實透明性與可解釋性」、「促進永續發展」等六大核心原則指引[114]，並在徵詢各界及專家學者意見後，就此六大原則為基礎，於 113 年 6 月 20 日發布《金融業運用人工智慧（AI）指引》，作為金融機構導入、使用及管理人工智慧的參考，也是我國第一份較為完整規範金融業使用人工智慧的政府文件。

系爭指引共分為總則及六大章節，總則包括人工智慧之定義、人工智慧生命週期、以「風險」考量之核心原則、風險評估時的考量因素及第三方業者監管等事項，第一章至第六章則是金融業在落實核心原則一至原則六時，依人工智慧之生命週期及相應可能產生的風險，所提出的指引重點及措施。

其中，涉及反歧視規範的依據位於對應核心原則二之第二章—重視公平性及以人為本的價值觀，其內容明訂：「（一）金融機構在使用 AI 系統之過程中，應儘可能避免演算法之偏見所造成的不公平。」此原則的目的是為了避免 AI 系統自動化的特性，而可能產生歧視或不公平的結果。其指出，「注意偏見（bias）之產生並儘可能避免歧視（discrimination)。」可以見得指引欲防免的是歧視的發生而非偏見，然而為了避免 AI 系統產生歧視，應定期審查、驗證模型，並注意偏見的產生以防免歧視的發生[115]。

在其定義的概念中，公平性意指金融機構運用人工智慧系統之決策不應對特定群體產生歧視結果，具體而言應符合「決策之合理性」及「決策之準確性及避免偏見」兩大面向。合理性的概念指出，並非一概禁止所有以個人屬性作決策的情形，若有合

---

[113] 金管會（2023/11/28），〈金管會就金融業運用人工智慧(AI)指引草案公開徵詢外界意見〉，https://www.fsc.gov.tw/ch/home.jsp?id=96&parentpath=0,2&mcustomize=news_view.jsp&dataserno=202312280004&dtable=News（最後瀏覽日：05/08/2024）。

[114]「金融業運用人工智慧(AI)指引」草案

[115] 金管會（2024），《金融業運用人工智慧指引》，頁 9。

doi:10.6342/NTU202500326

理的正當原因仍可容許，且為確保決策的準確及偏見最小化，應定期審查及驗證決策模型並檢驗相關資料，確保模型運算結果符合預期目的，而為了避免消費者受到 AI 系統的不利影響，金融機構應提供相關救濟管道[116]。指引亦提出公平性的落實方式，在「系統規劃及設計」階段應先確立金融機構規劃並使用 AI 系統的目的，並辨識出可能會受 AI 系統影響的不利群體以及可能被影響的程度，做成紀錄；在「資料蒐集及輸入」階段，首先應從資料面向檢視是否可能帶有偏見，例如從資料來源、蒐集方式，且需持續注意若帶有偏見，這些偏見是否會造成歧視。同時金融機構應注意使用多元的資料，避免依賴單一資料，以分散對特定群體的偏見風險，在使用有關「個人屬性」資料時，應特別注意目的和是否有使用的必要性，以及是否可能有其他替代屬性或替代方法；在「模型建立及驗證」階段，金融機構應先自行檢驗模型對於不同群體間的結果，確認該模型是公平無偏見的，若有偏見產生也應盡可能地調整，同時也應委外專業人員，定期審查及驗證，並將這些驗證結果留下紀錄，以供追溯及檢視結果。最後，在「系統部署及監控」階段中，金融機構仍應定期檢視、分析 AI 系統的結果是否具有歧視，若有歧視的發生應及時調整改善，且應辨識對受有不利影響的群體，是否係與 AI 系統的使用具有關聯性，若確實具有關聯性應降低對該不利影響群體使用 AI 系統。

第四節　小結

本章先從人工智慧與機器學習概述其定義，以及不透明性及公平性之問題，接著進而討論人工智慧信貸歧視的案例，從多起人工智慧模型應用在信貸服務中所產生的公平性疑慮，試圖將造成歧視的原因加以分類後，討論受保護特徵與非受保護特徵的歧視態樣。最後就我國現行法制下，梳理金融反歧視之相關規範，發現我國在公平信貸規範上的欠缺，而在金融機構開始使用人工智慧系統後，雖有如銀行公會頒布之《金融機構運用人工智慧技術作業規範》及金管會發布之《金融業運用人工智慧

[116] 金管會（2024），前揭註，頁 10。

doi:10.6342/NTU202500326

（AI）之核心原則》，對於公平性的要求已有一定的基礎，然而仍有不足之處。觀察目前的反歧視規範，可發現列舉出的受保護特徵多為歷史上受迫害的群體，立法者透過法律規範使歷史上的不公平不再重演。然而，若反歧視規範係建立在已發生不公平的受保護特徵上，則新型態的歧視原因恐將難以囊括其中。就此，本文認為反歧視規範的目的是追求平等、自由或是尊嚴，此應包括「預防」歧視帶來的不公平結果，亦有認為反歧視追求的是社會包容，使人們擁有自由、自治的權利，而有發展自身能力、追求生活的機會[117]。

[117] Hugh Collins, *Discrimination, equality and social inclusion*, THE MODERN LAW REVIEW, 66 (2003).

doi:10.6342/NTU202500326

第三章 公平信貸規範的理論與法制

前一章梳理了人工智慧歧視的發生以及其與受保護特徵的使用連結關係，特別是有意或無意透過與受保護特徵密切關聯的中性特徵，規避現行的反歧視規範，進而達到歧視目的的行為，是人工智慧歧視欲待解決的問題之一。從人工智慧運用在信貸服務中，亦發現確實造成公平性問題的歧視的案例，而梳理我國現行法制後，發現金融反歧視法制的欠缺，無法有效防免信貸服務中的歧視發生。因此，本章將參酌美國公平信貸規範，討論公平信貸規範的理論基礎與法制。

第一節　公平信貸規範之規範目的

在本節中，首先從公平信貸規範的規範目的及正當性出發，討論為何需要此規範介入私人間的借貸關係，在私人間契約自由及金融機構追求最大利益的自由下，為何應以法律規範金融機構，討論信用貸款對於人民而言的社會意義。

第一項　契約自由與公平信貸

第一款　契約自由與信用貸款中的正義

金融機構與借款人間的信用貸款關係乃民事上之契約關係，基於契約自由原則，金融機構原則上有追求其商業利益的自由，應享有決定是否與借款人締約的權利。然而現代社會生活亦有反歧視之普世原則，考量到信用貸款為社會生活中個人獲取資金的重要來源之一，故社會科學家和哲學家頗有認為債務（debt）為社會福利的一環，可能影響個人、兒童和家庭的生活而具有其社會公益性[118]。若個人因不可改變的先天或後天個人特徵而受有不公平的信貸程序，對其個人權利也是一種不正當的侵害。

第一目　民事關係中的反歧視規範

[118] John Linarelli, *Equality and Access to Credit: A Social Contract Framework*, LAW AND CONTEMPORARY PROBLEMS, 165 (2021), https://digitalcommons.tourolaw.edu/cgi/viewcontent.cgi?article=1789&context=scholarlyworks (last visited May 30, 2024).

doi:10.6342/NTU202500326

金融機構評估核貸金額與核貸決策時，係基於信用評估的預測結果，針對低信用評分亦即具有高信用風險的借款人，給予較低的貸款額度、較高的貸款利率或不予核貸，此為私法自治的當然結果。然而，若金融機構於信用評估預測時所使用的特徵或變數涉及男女、宗教、種族、階級、黨派等法定受保護特徵，因此做出不利於借款人的結果時，國家是否應介入管制此私法關係中的差別待遇？不無疑問。

根據國際公約及我國《憲法》第七條平等原則的規範[119]，國家不得以「男女、宗教、種族、階級、黨派...等」法定受保護特徵歧視個人，但憲法規定是否得直接適用於私人契約關係？憲法基本權的規範主體原應為國家與人民間的關係，至於人民與人民間的私法關係中是否得直接援引平等權而主張不受歧視的權利？涉及基本權第三人效力的爭議。依目前我國多數學說見解，應採取間接適用說[120]，即基本權不直接適用於私法關係。儘管如此，多數見解亦認為個人仍可援用民法上公序良俗及誠信原則等概括條款，透過制度性保障的解釋，由法官針對具體案件判斷須優先保護的法益，落實基本權的保障，故在私法自治的契約關係中，也應有憲法原則的適用，此為私法的憲法化。

我國學界過往亦曾有討論透過民法第 18 條第 1 項規定於私法關係中保障個人平等權的可能性，平等權的保障既係維護人性尊嚴，為人格發展的重要權利之一，故在私法關係中如個人的平等權被他人所侵害，是否得援引人格權行使其排除權利？就此國內有學者曾認為，本條之人格權「侵害」具有高度不確定性及抽象性，若肯認人格權可適用於歧視行為，契約自由原則可能被架空，故解釋上似不宜過寬[121]。

---

[119] 如：《公民權利和政治權利國際公約》、《經濟、社會、文化權利國際公約》

[120] 葉啟洲（2014），〈民事交易關係上之反歧視原則－德國一般平等待遇法之借鏡〉，《東吳法律學報》，26 期，頁 40；王澤鑑，法律思維與民法實例，自版，頁 241-242，1999 年 10 月。；楊舒嵐（2010），《憲法價值於私法關係之體現─以隱私權與表現自由之衝突與調和為例》，國立臺灣大學碩士論文， https://doi.org/10.6342/NTU.2010.10595；許志雄，人權的私法間效力論，載：人權論──現代與近代的交會，元照，頁 114，2016 年 1 月。

doi:10.6342/NTU202500326

本文認為，平等原則適用於私法關係時，仍須考量與契約自由間的價值權衡，不應過分偏袒一方。反歧視之平等原則固然是普世人權的實現原則之一，但契約自由亦是落實自我實現的途徑，也是我國憲法第 22 條保障的概括基本權[122]。民法的概括規定雖可將憲法原則引入私法契約中，然而面對契約自由的價值平衡時，仍不足以作為明確的行為規範。因此，本文認為為權衡信貸歧視議題中的公平信貸與契約自由兩項原則時，應有明確的法律規範訂定，以符合憲法第 23 條法律保留原則與比例原則[123]。

第二目　反歧視規範的理論基礎

信用貸款給予人民獲取資金的管道和機會，個人透過貸款得以取得較原先預期的工作收入、生活所得以外更多的金錢，對於其經濟生活將有更多的選擇。因此，有論者從社會學及跨學科的角度觀察，認為信貸取得（access to credit）是社會基本結構的一部分，其也涉及了社會資源的有效分配[124]。

在信用貸款的徵信階段，金融機構透過信用評分、風險評估、貸款與收入之比例等指標，以評估是否為合適的核貸對象。信貸獲取機會的正義，取決於人民是否受到金融機構公平的信貸審核過程，且基於社會的整體公正性受有核貸的機會，其背後的「公正性」實涉及國家的整體福利政策價值觀[125]。若將信貸取得作為一種社會基本利益，縱使借款人在還款上可能有較高的貸款回收風險，也不能因此限制其獲取信貸的機會[126]。但如採此政策觀，金融機構提供信貸的成本自然將大幅提升，故需仰賴國家公權力的介入，才有可能實現信貸作為社會福利的部分，為人民提供具有正義的信貸服務。本文認為，在兼顧金融機構合理追求商業利益的權利下，一般信貸中的反歧視規範追求的並非是將信貸作為社會福利而要求金融機構完全不得以受保護特徵考量授

[122] 釋字 576 號解釋意旨
[123] 葉啟洲，前揭註 120，頁 44。
[124] Linarelli, *supra* note 118, at 175.
[125] *Id*. at 180.
[126] *Id.* at 179.

doi:10.6342/NTU202500326

信風險，此亦非金融機構所應負責的義務，因此信貸反歧視所追求的僅是避免階級上的不平等影響個人所接觸的信貸程序。

綜上，信貸服務中所追求的公平信貸，其目的在於對具有受保護特徵的一定群體的分配正義，金融機構對個人進行信用風險評估時，縱使該受保護特徵可能與信用風險的提高有正相關性，不得因個人具有受保護特徵而予以不利益的結果。此政策固然可能一定程度地犧牲金融機構的商業利益，但是追求社會分配正義所必要。因此，在兼顧金融機構的締約自由及人民的平等權保障之下，信貸公平與反歧視規範追求的目標，應是金融機構不得將特定的法定受保護特徵列為影響信用評分的變數，使具備該特徵的借款人可獲取起碼平等的評估程序，但在此基礎上，金融機構仍可依據借款人的其他個人條件評估其信用風險決定是否核貸。

第二款　公平、平等與正義

肯認國家規範金融機構之信貸決策不得有歧視行為的正當性後，為建立相應的立法規範或監管措施，需釐清此舉追求的具體目標為何。本文以下先從公平、平等、正義三者的區別談起，進而探討信貸反歧視規範的核心目標。

第一目　平等（Equality）

平等係指對群體中所有人平均分配資源，不考慮個別需求或能力差異等個人背景，皆給予相同的資源或機會，平等強調相同標準的一致性和無差別待遇[127]。例如美國獨立宣言所示：「人人生而平等，造物主賦予他們不可剝奪的權利，包括生命、自由和追求幸福的權利。[128]」

[127] Georgetown Online Public Health, *Equity vs. Equality: What's the Difference?* (November 5, 2020), https://onlinepublichealth.gwu.edu/resources/equity-vs-equality/. (last visited Jan 1, 2025); World Economic Forum, *International Women's Day: What's the difference between equity and equality?* (Mar 3, 2023), https://www.weforum.org/stories/2023/03/equity-equality-women-iwd/ (last visited Jan 1, 2025).

[128] 原文：All men are created equal, that they are endowed by their Creator with certain unalienable rights, that they are among these are life, liberty and the pursuit of happiness.

doi:10.6342/NTU202500326

第二目　公平（Equity）

公平係指對不同個體間的合理對待。除了平等分配資源外，亦會考慮人與人之間的差異性，並強調過程及結果的公正性，根據具體需求給予適當的對待，以確保資源分配的合理性[129]。

在公平的脈絡下，資源分配須依據個體間的差異或特定的具體需求，而有不同的標準或應對，其著重的重點在結果的公平性，而非表面上的齊頭式平等[130]。例如教育資源分配應考慮城鄉差距、資訊落差或社會經濟地位的差異，根據具體不同的需求而給予相應的幫助，確保每個學童皆有平等的學習權。

第三目　正義（Justice）

正義的概念是價值選擇的結果，難以對其給出一個普世認同的概念，因為其可能涉及法律、道德和倫理等內涵。在常見的正義理論中，功利主義代表邊沁（Jeremy Bentham）提出，正義係指追求社會整體的最大幸福值，並以結果評斷對與錯，其認為「效用」應優先於「公平」，此為效用論（Utilitarianism）[131]。

自由主義的代表為約翰・羅爾斯（John Rawls），其在《正義論》（A Theory of Justice）中批判了邊沁的效用論和馬克思主義，其主張在追求效率最大化的同時，每位公民仍應有同等的自由和公平的機會去追求理想，其中有兩大原則[132]：第一原則為，每個人都享有平等、自由的基本權利；第二原則為差異原則與機會平等原則，前

[129] *Id.*
[130] *Id.*
[131] Jereny Bentham, *An Introduction to the Principles of Morals and Legislation* (revised 2017), https://www.earlymoderntexts.com/assets/pdfs/bentham1780.pdf (last visited Jan 3, 2025)
[132] John Rawls, *A Theory of Justice*, HARVARD UNIVERSITY PRESS (1971).〈平等、公平、現實、正義(解放)之間的差異〉，https://medium.com/vincent-chen/%E5%B9%B3%E7%AD%89-%E5%85%AC%E5%B9%B3-%E7%8F%BE%E5%AF%A6-%E6%AD%A3%E7%BE%A9-%E8%A7%A3%E6%94%BE-%E4%B9%8B%E9%96%93%E7%9A%84%E5%B7%AE%E7%95%B0-79f7043e8d6e（最後瀏覽日：01/06/2025）。

doi:10.6342/NTU202500326

者係社會應優先考慮弱勢成員的利益，後者則為所有人都應該擁有公平機會競爭各種職位或地位。

最後，社群主義以邁可・桑德爾（Michael Sandel）代表，其著作《正義，一場思辨之旅》探討了社會正義問題，他認為正義係社會或制度的整體公正性，應以公共精神為基礎，包括分配正義（資源和機會的公平分配）、程序正義（決策過程的公正性）和矯正正義（對不公正行為的矯正和懲罰）[133]。

平等係為人們提供相同的資源，公平是指為人們提供個人達成目標所需的資源，而正義指的是為人們提供手段，讓他們有相同的機會實現自己的目標[134]。平等和公平可以透過公平的演算法和方法來解決，而正義則需要外部代理（outside agent），例如平等信用法。舉例來說，在大學入學名額有限的情形下，平等的手段是增加招生人數，使人人都有機會就讀；公平的手段是例如給予經濟困難的學生一定的配額或給予獎學金，而促進正義的手段可能是制定政策投入於高中教育，使高中畢業生都有機會得以進入大學，平等、公平和正義間最大的區別在於正義需要仰賴外部的力量才得以達成，正義需要仰賴決策者採取行動才得以達成，而公平的演算法可以實現平等及公平[135]。

金融反歧視規範的核心宗旨即為平等、公平或正義上的價值取捨，又可謂平等主義（Egalitarian）、自由主義（Libertarian）和尊嚴主義（Dignitarian）的選擇。平等主義追求人生而為平等，應有受到公平待遇的權利；自由主義認為規範應促進自由、自治；尊嚴主義則是追求人性尊嚴的保障[136]。其中平等主義在歐洲啟蒙運動時被大力倡導，如今已是普世人權基本原則，透過反歧視法規範追求平等待遇（直接歧視）或實

[133] Sandel, Michael J., *Justice: What's the Right Thing to Do?*, FARRAR STRAUS & GIROUX (2010).
[134] Ana Cristina Bicharra Garcia, Marcio Gomes Pinto Garcia, and Roberto Rigobon, *supra* note 49, at 2.
[135] *Id.*
[136] Wachter, *supra* note 17, at 34.

doi:10.6342/NTU202500326

質平等（間接歧視），試圖終結不公平的壓迫[137]。因此，金融反歧視法之規範目的，不僅只在於追求平等或公平，更是需要國家介入的外部力量已達到正義的追求。

第二項　傳統信貸審核與公平信貸

梳理公平信貸規範所追求的目的後，接著則需觀察傳統以人力審核信貸時，有何審核上遵循的標準及方法，且從這樣的審核過程中，為何會產生歧視？而現行公平信貸規範又該如何規範這樣的歧視。

在傳統的信貸評估中，金融機構係以「風險」為基礎，計算借款人的違約機率、核貸的機會成本、未來的貸款回收率等[138]。而風險的計算，係透過金融機構對借貸人的徵信與授信審核，其涉及金融機構如何以風險為基礎以考慮借貸人的個人條件，以評估其履行貸款義務之能力及給予如何的貸款條件[139]。在傳統人力的信貸審核中，金融機構主要以信用報告及信用評分兩種方法以評估借貸人的風險及決定貸款定價，信用報告係由第三方機構發布關於借貸人的信用及付款歷史之信用度[140]，針對個人及企業的信用評估[141]；金融機構亦常參考源自美國的 5C 理論及 5P 理論為建立信用評估方法之標準[142]，這些標準也是建立信用評分系統時的核心原則，為金融機構提供系統化的評估方法，確保貸款決策的合理性和穩健性。以下簡述金融機構進行信貸評估決策的遵循原則與方法。

第一款　金融機構信貸決策的標準

第一目　5C 理論[143]

---

[137] *Id*. at 35.
[138] Wu, *supra* note 4, at 73.
[139] Wu, *supra* note 4, at 74.
[140] 如 Equifax、Experian 和 TransUnion 等消費者信用報告機構出具之信用報告
[141] Wu, *supra* note 4, at 74.
[142] 康建民（2008），《個人貸款信用評分模型-以儲蓄互助社為例》，頁 12，國立暨南大學財務金融學系碩士論文。
[143]New York Institute of Finance, *The C's and P's of Credit Risk Analysis,* https://info.nyif.com/do-know-your-cs-and-ps-for-credit-analysis/ (last visited Jan 6, 2025).

doi:10.6342/NTU202500326

所謂 5C 理論，係指金融機構做成信貸決策時，應考量以下五個 C 開頭的面向：

一、Conditions（條件）

條件指的是影響貸款的外部因素，如經濟環境、產業情勢、借貸目的和貸款條件等。金融機構會將當前的經濟形勢、政策變動和市場風險等外部因素納入評估，全方面了解外部影響條件，以確保貸款決策的合理與風險得以控制。

二、Collateral（擔保品）

擔保品係指借貸方提供的擔保財產，目的在擔保借貸方無法在期限內還款時，金融機構可以處置擔保品以填補虧損。常見的擔保品包括不動產、汽車和其他有形資產，金融機構須評估擔保品之價值和可變現性，以確保借貸方違約時能有效減少損失。

三、Character（品德）

品德指的是借貸方的信用歷史和個人聲譽，包括其過去的還款記錄、財務紀律和誠信度。金融機構通常會查閱信用報告以了解借款人的信用評分、過去的貸款紀錄和是否有拖欠款項的歷史。評估品德的目的是了解借款人的還款意願，確保其具備履行貸款義務的誠信。

四、Capacity（償債能力）

償債能力指的是借貸方的還款能力，金融機構通常通過分析借貸方的收入、支出、現金流和負債水平來衡量。金融機構會檢查借貸方的收入來源、職業穩定性和現有債務比率等指標。評估償債能力的目的旨在確保借款人有足夠的財力來按時還款，減少貸款風險。

五、Capital（資本）

doi:10.6342/NTU202500326

資本指的是借款人的淨資產，即其擁有的資產減去負債後的淨值。這一要素於企業貸款中特別重要，因為它代表著借款人應對財務困難的緩衝能力。評估資本的目的在於了解借款人在經濟困難時期的償債保障。

第二目　5P 理論[144]

所謂 5P 理論，係指金融機構做成信貸決策時，應考量以下五個 P 開頭的面向：

一、Purpose（目的）

目的指的是借貸方申請貸款的用途，如擴大生產、技術研發或市場拓展等。明確的貸款用途能增加貸款的透明度和可控性。評估貸款目的有助於了解借貸方的發展規劃和資金使用的合理性。

二、Payment（還款）

還款指的是借貸方的償債計劃和能力，包括對借貸方的現金流預測和財務規劃。了解借貸方的收入來源和現金流狀況，是金融機構評估其還款能力的核心。確保借貸方具備足夠的還款能力和可靠的償債計劃，降低貸款風險。

三、Protection（保護）

保護指的是貸款的保障措施，如擔保品或保險等。提供有效的擔保和風險緩釋工具，是保障金融機構貸款安全的重要手段，確保在違約情況下能有效降低損失。

四、Person（人員）

人員指的是借貸方的管理團隊和員工，包括其經驗、能力和專業素質。評估管理團隊的領導能力和決策水平，是了解借貸方穩定性和發展潛力的關鍵。良好的管理團隊能有效推動企業發展，降低經營風險。

[144] *Id.*

doi:10.6342/NTU202500326

五、Perspective（前景）

前景指的是企業的未來發展潛力和市場前景，包括其行業地位、競爭力和市場趨勢。分析市場需求和競爭環境，預測企業的發展前景。確保企業具備良好的市場前景和持續增長能力，增加貸款的可行性和安全性。

第二款　常見信用評估方法

金融機構針對信用風險常見的評估方式有經驗法則、信用評等法、信用評分法、混合制、統計方法與資料驅動法等[145]，金融機構會根據其運營狀況、授信政策、目標客群與營運成本的不同，結合數個方法以不同的權重比例評估。以下簡介各評估方式的特點[146]：

第一目　經驗法則（Rules of Thumb）

經驗法則主要仰賴授信人員的過去經驗和主觀判斷決定是否核貸。運用此方法在執行上較為容易，且具有較高的彈性，但易流於個人主觀價值判斷而缺乏客觀性，也因缺乏統一客觀標準，可能產生同一案件在不同人員審核評估下有不同的結果，容易引發弊端和不公正情形[147]。

第二目　信用評等法（Credit Grading System）

信用評等法是將經驗法則轉化為制度化程序，將客戶品質劃分為不同等級。在評估申貸客戶時，根據各等級的定義或條件進行綜合評價，並給予相應的等級。此方法旨在減少授信人員主觀判斷的影響，但最終的評等仍是由授信人員主觀判斷。此外，評等項目的選擇和標準的確定較為困難，若定義不夠明確，即容易產生評估差異[148]。

---

[145] 李婷婷（2019），《小額消費信貸信用評分模型的建置–以台灣某銀行為例》，頁 3，國立清華大學財務金融碩士在職專班碩士論文。
[146] 以下簡介整理自康建民，前揭註 147，頁 19。
[147] 康建民，前揭註 104，頁 17。
[148] 康建民，前揭註 104，頁 18。

doi:10.6342/NTU202500326

第三目　信用評分法（Credit Scoring System）

信用評分法將評估項目量化，制定各項變數的具體分數，形成最終的信用評分表。得分越高，代表償債能力越高。此方法的優點在於其內容簡易且評分過程客觀，但評分項目的選擇和權重設定相對複雜。目前，各大銀行在消費金融業務中廣泛使用信用評分制度，能有效應對業務要求的時效性和大量客戶的需求[149]。

第四目　混合制

混合制結合了信用評分法和信用評等法，根據最終得分進行分級，並給予相對應的授信條件。混合制的方法結合了兩種方法的優點，提高了評估的準確性和一致性，但也增加了其複雜性，需在實施過程中保持兩種方法的平衡[150]。

第五目　統計方法

金融機構也可利用例如 Logistic 迴歸分析、區別分析和 Probit 模型等，透過資料分析建立信用評估模型。此方法具有較高的客觀性和科學性，評估結果更具說服力和可操作性。然而，模型構建和應用較為複雜，需要大量資料運行[151]。

第六目　資料驅動法：

資料驅動法利用電腦資訊技術如人工智慧，通過大量歷史資料進行訓練，建立自動評核系統。此方法高度自動化且具有客觀性，但其演算法運作原理複雜且難以解釋，研究者通常只能依照操作規範進行操作，不完全了解內部運作，存在相對的應用風險[152]。

第二節　公平信貸與反歧視規範的內容

[149] 康建民，前揭註 104，頁 18。
[150] 康建民，前揭註 104，頁 18。
[151] 康建民，前揭註 104，頁 18。
[152] 康建民，前揭註 104，頁 18。

doi:10.6342/NTU202500326

第一項　反歧視規範的立法模式

根據歸納，現行的反歧視規範依據法條文字的結構組成，可區分為三種：封閉型、開放型以及混合型[153]。

第一款　封閉型

封閉型系統係訂定詳細的歧視定義，列舉所有禁止歧視的受保護特徵及相應允許歧視的正當理由[154]。封閉型多用於有具體特定的規範對象，例如歐盟有許多指令即是採用此方式，如歐盟消費者信貸指令明確禁止歧視性借貸，禁止對於合法居住在歐盟消費者基於「國籍、居住地、性別、種族、膚色、民族或社會出身、遺傳特徵、語言、宗教、信仰、政治或任何其他見解、少數民族成員、財產、出生、殘障、年齡或性取向」的歧視[155]。

第二款　開放型

開放型系統多半不會在規範中明訂具體、個別的受保護特徵，僅敘明抽象的規範目的、價值，例如《歐盟基本權利憲章》第 20 條：「法律之前，人人平等。[156]」以開放形式規範平等權[157]，相較封閉型的規範方式，開放型系統可納管任何形式的不公平情形，因此，應用此類開放性系統的法規範，法院可以合理考量任何相關的個人、群體或情況之間的差異，或任何其他可能證明差別待遇的原因[158]。

第三款　混合型

---

[153] Gerards and Zuiderveen Borgesius, *supra* note 18, at 11.
[154] Aalt Willem Heringa, *Standards of Review for Discrimination: The Scope of Review by the Courts*, NON-DISCRIMINATION LAW: COMPARATIVE PERSPECTIVES 25, 27 (Titia Loenen & Peter R. Rodrigues eds., Kluwer Law International 1999).
[155] Art. 6 of the proposal, see already recital (45) of the current Consumer Credit Directive 2008/48/EC.
[156] Article 20: Everyone is equal before the law.
[157] Gerards and Zuiderveen Borgesius, *supra* note 18, at 27.
[158] *Id.* at 28.

doi:10.6342/NTU202500326

混合型系統結合封閉型與開放型的特點，除了規範具體明確的受保護特徵外，也有允許歧視的正當理由的解釋空間。封閉型的受保護特徵多針對特定的標的，以列舉方式訂定，而混合型則是以例示的方式，將具有相同規範精神的受保護特徵列舉後，仍保有其他相同類型特徵的可能性。

例如歐盟《歐盟基本權利憲章》第 21 條：「不得基於任何理由歧視，例如基於性別、種族、膚色、血源或社會背景、面容外貌、語言、宗教與信念、政治或任何其他意見、少數族裔成員、財產、出生、殘障、年齡或性傾向等之理由，均應被禁止。」，此類型的規範方式多會出現「不得以任何理由進行歧視，例如...」的概括加例示的字句。傳統的反歧視規範多採用封閉型或混合型立法形式，將受保護特徵正面列舉於法律條文中，明示禁止歧視的受保護特徵，但又同時彰顯規範精神而具有象徵性的意義。

歧視是種複雜的現象，其背後的原因可能是多元交錯的。明定受保護特徵也是彰顯立法政策的價值選擇，反映當下的社會脈絡下所重視的價值體現。透過具體明定受保護的特徵，反歧視法規得以規範基於主觀品味的直接歧視（例如基於傳統的性別或種族主觀偏見而拒貸），也得以進一步規範缺乏主觀要件的間接歧視[159]。

本文認為，三種立法型態皆有各自的優缺點，封閉型可以清楚具體指出受規範的保護特徵，具有明確性方便遵循，但缺點則是其規範的範圍固定，導致其面對新型態歧視的應變能力不足。為確保反歧視規範的有效性，本文認為公平信貸規範應採取偏向開放型的混合型立法，避免金融機構利用表面中性的特徵替代受保護特徵規避公平信貸規範。另一方面，開放型立法在面臨新型態的歧視時固然具有靈活性，但同時也容易流於抽象、不精確，甚至有違反法明確性的可能，故在開放型立法中加入例示的

[159] Gerards and Zuiderveen Borgesius, *supra* note 18, at 33

doi:10.6342/NTU202500326

保護特徵，亦即採取混合型立法，較有助於增加規範的明確性而有遵循的依據，同時仍保有解釋的開放空間。

第二項　反歧視規範下的歧視態樣

具體是用反歧視規範時，需進一步釐清此類規範所欲規範的歧視態樣。美國法下的文獻上多將歧視進一步區別為差別對待歧視（disparate treatment）及差別影響歧視（disparate impact）討論[160]，以下分別就其定義及成立要件討論之。

第一款　美國法下的差別對待歧視與差別影響歧視

第一目　差別對待歧視

差別對待歧視是指針對受保護特徵所保護的人群，以故意、直接且顯著的方式進行歧視。典型的案例是針對種族、膚色、宗教、性別或國籍等法定的受保護特徵做成差別待遇，例如在 Monson v. Rochester Athletic Club 案中[161]，法院認為差別對待歧視是指「基於種族、膚色、宗教、性別或國籍而給予不如其他人的待遇」；在 DeJung v. Superior Court 一案中，法院對差別對待歧視的定義為「基於禁止的理由的故意歧視」[162]。

第二目　差別影響歧視

差別影響歧視是指表面上不造成歧視、看似公平，但實際上卻會對具有特定受保護特徵的族群產生不利影響的結果。例如本文上述的美國房屋抵押貸款中銀行的「劃紅線」拒絕核貸行為[163]，銀行以申貸人居住於紅線區域的社區內為理由拒絕提供貸款，雖然「劃紅線」的行為本身並不違法，然而實際上以紅線區域內的居民為區分的

---

[160] Ana Cristina Bicharra Garcia, Marcio Gomes Pinto Garcia, and Roberto Rigobon, *supra* note 49, at 2; Katja Langenbucher, *supra* note 10, at 10, 23; Wu, *supra* note 4, at 94.
[161] *Monson v. Rochester Athletic Club*, 759 N.W.2d 60 (Minn. Ct. App. 2009)
[162] *DeJung v. Superior Court*, 169 Cal.App.4th 533, 87 Cal. Rptr. 3d 99 (Cal. Ct. App. 2008)
[163] Investopedia, *Effects Test: What It Means, How It Works* (2022), https://www.investopedia.com/terms/e/effects-test.asp (last visited Jan 4, 2025).

doi:10.6342/NTU202500326

原因，是因為該區域內多居住著同一種族的居民，所以表面上畫紅線的區分行為，實質上是為了對該特定種族的人民為拒絕核貸的不利影響，因此產生差別待遇。

差別影響歧視的概念於美國法下最早係於 1971 年最高法院 Griggs v. Duke Power Co.案中獲得確認[164]。本案為一就業歧視案件，於該案中，最高法院認為受歧視的受雇人不需要證明行為人的「故意歧視意圖」，而僅需證明其雇用措施（Employment practice）會對特定受保護群體造成差別影響，即可證明行為人構成歧視；但法院同時也允許行為人舉證抗辯其行為的正當合理性，以免除相關責任[165]。

差別影響歧視雖於本案以後獲得被承認，並被編入民權法第七編的 Section 703(a)(2)而有法律上的依據[166]，但因缺乏明確的要件與舉證責任規範，在後續的訴訟中仍有不少困擾[167]。故美國於 1991 年修訂民權法時，除了將差別影響歧視正式列入第七編中的歧視類型之外[168]，也明定舉證責任之分配，使訴訟上的標準更加明確。

上述的差別影響歧視在民權法第七編的規範脈絡下，免除了行為人需有主觀上的歧視意圖要件，僅需達到客觀上的差別影響程度即可成立，有論者稱此為客觀理論的歧視或過失歧視[169]。欲成立客觀理論的差別影響歧視，主要係以統計上結果差異，也就是不以主觀的歧視故意為要件認定是否具有歧視，而係以結果的統計差異來認定歧視的存在[170]。

有鑒於除了故意歧視之外，歧視仍包含許多不同態樣、難以察覺或結構性的偏見及刻板印象，且其成因亦可能是非故意（unintended）、無意識（unconscious）或潛意

---

[164] *Griggs v. Duke Power Co.*, 401 U.S. 424, 91 S. Ct. 849 (1971)
[165] *Id*, at 431.
[166] 42 U.S. Code § 2000e–2
[167] 黃昭元（2012），〈論差別影響歧視與差別對待歧視之關係－評美國最高法院 Ricci v. DeStefano(2009)判決〉，《中研院法學期刊》，11 期，頁 25。
[168] 42 U.S. Code § 2000e–2 (k)
[169] 黃昭元，前揭註 167，頁 45; George Rutherglen, *Disparate Impact, Discrimination, and the Essentially Contested Concept of Equality*, 74 FORDHAM L. REV. 2313 (2006);
[170] 黃昭元，前揭註 167 ，頁 52。

doi:10.6342/NTU202500326

識（subliminal）下形成的[171]。因此，本文認為在現代生活中的平等理論所欲規範的歧視，亦應包含這些潛在且未被察覺的偏見、歧視，故將差別影響理論與平等理論相連結，應可達到更佳的規範目的。

第二款　歐盟法下之直接歧視與間接歧視

與美國法不同，歐盟法則是將歧視態樣區分為直接歧視和間接歧視。直接歧視可將意圖分為顯性及隱性，顯性之直接歧視是最直觀且外顯的歧視，通常以常見的受保護特徵形態出現，例如直接以性別作為給予不利貸款條件的因素；而隱性的直接歧視則是具有歧視之意圖，只是行為人有意遮掩其歧視行為，由於其主觀上仍有歧視故意，仍為直接歧視的類型，例如金融機構要求借款人申請信貸時須提供長期居住證明，且必須為本國出生人民，這樣的限制可能移民或某些少數族裔有不利的影響，但表面上看似對所以借款人都要求一樣的提供文件。

相對於直接歧視，間接歧視的概念源自於美國法 Griggs 案確立之差別影響歧視，後由英國法採用後，在歐陸國家將其稱為間接歧視[172]。間接歧視與直接歧視最大的差別在於，間接歧視在意的是結構性不利益的因素，像是特定規範或措施所產生的歧視效果，且間接歧視著重於特定群體的差別待遇，因此，間接歧視的概念是要以規範禁止與直接歧視具有類似歧視效果的情形[173]。

禁止歧視原則是《歐盟基本權利憲章》（Charter of Fundamental Rights of the European Union）中重要原則之一[174]，該憲章第 21 條規定：「1. 任何基於如同性別、種族、膚色、血源或社會背景、面容外貌、語言、宗教與信念、政治或任何其他意

---

[171] Ann C. McGinley, *Discrimination Redefined*, 75 MO. L. REV 443, 448-49 (2010) ; Charles R. Lawrence III, *The Id, the Ego, and Equal Protection: Reckoning with Unconscious Racism*, 39 STAN. L. REV. 317 (1987)；黃昭元，前揭註 167，頁 53。

[172] Sandra Fredman, *Discrimination Law*, Oxford University Press, 178 (2011).

[173] 陳靜慧（2017），〈歐洲人權法院及歐洲法院對間接歧視之適用〉，《憲法解釋之理論與實務》，9 期，頁 395。

[174] 憲章第 21 條第 1 項

doi:10.6342/NTU202500326

見、少數族裔成員、財產、出生、殘障、年齡或性傾向等之理由歧視，均應被禁止。2. 在適用《歐洲共同體條約》與《歐盟條約》的範圍內，任何基於國籍之歧視均應被禁止。」在憲章規範前，禁止歧視原則係以判例法的型態出現，被認定為一般法律原則[175]，間接歧視的禁止雖未直接被規範在憲章中，但一般而言也被包含在禁止歧視原則之保護範圍內[176]。

第三項　歧視的例外正當化事由

基於反歧視的普世原則及法律對受保護特徵的規範，凡是基於受法律明定列舉之保護特徵進行差別待遇之行為，均應受到禁止，然而於特定情形，政府或個人仍然可能基於追求合法利益之正當理由而對某些群體進行區分，例如警察招考因工作性質之業務上正當性而要求應考者有身高上的限制[177]。因此，即使考量人權保障，受保護特徵亦非絕對不可侵犯，若符合比例原則的檢驗且具有正當理由時，基本權仍有退讓的可能。

然而追求合法利益的不平等行為容易受到歷史背景導致的系統性歧視影響，無意間強化對某些受保護群體的負面刻板印象，例如在南非女性難民聯盟案件中，政府以難民的可靠性有疑慮為由，禁止難民從事私人保全行業。南非法院認為此舉即係基於刻板印象的歧視[178]，並強調了在追求合法利益的同時，必須考慮到不平等待遇的潛在後果。

在一些反歧視法的規範下[179]，如美國民權法第七章的就業歧視禁止，在未使用受保護分類標準的差別影響歧視中，只要決策係追求合法正當的利益，且其無法透過其他較少歧視的替代方案達成目的時，通常即可被認為其有正當商業需求，從而可豁免

[175] 陳靜慧，前揭註 173，頁 391。
[176] 陳靜慧，前揭註 173，頁 392。
[177] Bar-Gill, Sunstein, and Talgam-Cohen, *supra* note 52, at 30.
[178] *Union of Refugee Women v. Director, Private Security Industry Regulatory Authority* (CCT 39/06) [2006] Z.A.C.C. 23, paras 122–123 (S. Afr.).
[179] 如 ECOA、民權法第七章

doi:10.6342/NTU202500326

歧視責任。在法院判決中，對於正當理由的檢驗通常包括「歧視行為是否到正當、重要且達到合理的目標」及「正當理由與被質疑的歧視行為間具有明顯可驗證的關聯性」[180]。而在信貸案例中，法院在評斷信貸歧視案件時，若金融機構並未直接故意使用受保護分類標準做成貸款決策時，多半認為金融機構為了定價而評估借款人信用風險的需求屬於「正當商業需求」（Legitimate Business Necessity），若其採取的分類標準有助於預測相關結果（例如計算債權回收和貸款違約風險），即被認為有助於滿足正當商業需求[181]。

第三節　比較法下的公平信貸

第一項　公平信貸規範

美國法下針對信貸歧視的規範主要以《平等信貸機會法》（Equal Credit Opportunity Act, ECOA）[182]和《公平住居法》（Fair Housing Act, FHA）為主，ECOA係針對信用貸款上的歧視，而 FHA 則是針對房屋抵押貸款的歧視。以下分述之。

第一款　ECOA 與 Regulation B

ECOA 於 1974 年頒布，訂於《美國法典》（United States Code）第 15 編第 1691 條至第 1691f 條。ECOA 第 1691a 條明定，授信方在信用交易中的任何環節中「基於種族、膚色、宗教、國籍、性別或婚姻狀況或年齡」而歧視申請者的行為，均為違法行為[183]。根據 ECOA 的授權，CFPB 進一步頒布相關施行規則[184]，並以 Regulation B 作為 ECOA 的配套措施[185]，明列實質上與程序上的規範。

---

[180] U.S. Department of Justice, *Federal Coordination and Compliance Title VI Manual Section VII- Proving Discrimination- Disparate Impact*, https://www.justice.gov/crt/fcs/T6Manual7#P. (last visited Nov 10, 2024).
[181] See, e.g., *A.B. & S. Auto Service, Inc., v. South Shore Bank of Chicago*, 962 F.Supp. 1056, 1061 (N.D. Ill. 1997); *Lewis v. ACB Bus. Services, Inc.*, 135 F.3d 389, 406 (6th Cir. 1998).
[182] 15 U.S.C. §1691 (2022).
[183] 15 U.S.C. §1691.
[184] The Equal Credit Opportunity Act, https://www.justice.gov/crt/equal-credit-opportunity-act-3
[185] Regulation B (Reg. B), 12 C.F.R. pt. 202 (2007).

doi:10.6342/NTU202500326

Regulation B 原係由美國聯邦儲備委員會（Federal Reserve Board, FRB）頒布，訂於《美國聯邦法規》（Code of Federal Regulations）第 12 編第 202 條以下。為促進信用良好的申請者得以公平地申請信貸，而不因其他個人因素而受到歧視，其第 202.5 條規範債權人不得詢問申請人或與信貸交易有關的任何其他人的種族、膚色、宗教、國籍或性別等資訊，盡可能減少歧視發生的可能[186]。

在第 202.6 條 a 項中，則訂定了消費者一般資料使用規範，明定資料的使用不可用來歧視申請者，該項的註腳中更提及：「根據其立法沿革，國會在此規範中有意採用效果測試（Effects Test），如 Griggs 案及 Albemarle Paper 案中判斷歧視的構成[187]。」如前述，Griggs 確立了差別影響歧視，其判斷歧視的發生係採用「統計上顯著」（statistically significant）的判斷標準，進行效果測試檢驗個別政策或行為是否對受保護群體產生了歧視性影響，且不問行為人是否具有主觀意圖為之。該測試會檢驗基於中性特徵做成的政策或行為對不同受保護群體的影響，進而評估其是否對受保護群體產生了不成比例的影響而構成偏見或歧視[188]。

Regulation B 第 202.6 條明文禁止差別影響歧視，但是否代表其母法 ECOA 亦採此立場，於美國有所爭論。有論者認為，根據 ECOA 本文的規範，其僅禁止根據列舉之受保護特徵歧視借款人，如第 1691(a)(1)條禁止基於種族、膚色、宗教、國籍、性別或婚姻狀況或年齡而有信用交易上的歧視，故難以認為 ECOA 承認差別影響歧視並採用效果測試[189]。但亦有學者認為 ECOA 雖然並未明文包含差別影響歧視的概念，然而該法中仍有其他規範得以推測其實際上承認差別影響歧視，例如第 1691(b)(3)條即規定信

[186] 12 C.F.R. §202.5 (b)
[187] *Albemarle Paper Co.* v. *Moody*, 422 U.S. 405 (1975)
[188] 黃昭元，前揭註 167，頁 27。
[189] Winnie F. Taylor, *The ECOA and Disparate Impact Theory: A Historical Perspective*, 26 J. L. & POL'Y 575, 588 (2018), https://brooklynworks.brooklaw.edu/jlp/vol26/iss2/3 (last visited Jan 1, 2025); *Inclusive Cmty. Project, Inc.*, 576 U.S. at 534 -38.

doi:10.6342/NTU202500326

貸評分系統需對年齡無負面影響方得使用[190]，可見其亦考慮表面中立但可能產生負面影響的狀況[191]。

本文認為，綜合 ECOA 中其他條項規範的隱含意義，加上 Regulation B 已明文填補該法的規範細節，故整體觀察而判斷該法的立法目的後，應認為 ECOA 亦規範表面中立但有實質影響的差別影響歧視。就此亦有學者綜合 ECOA 與 Regulation B 的立法沿革與條文內容綜合觀察之後，認為 FRB 對於效果測試在信貸交易中的解釋為：「若使用特定資訊於核貸時，縱使該資訊未在明文禁止之受保護特徵範圍內，但如使用該資訊而產生拒絕向受保護群體核貸的不利影響時，亦可能違法，除非借貸方能證明該資訊與信用可靠度具有明顯關係。」[192]換言之亦承認差別影響歧視受其規範[193]。

2010 年美國國會通過 Dodd-Frank Act，其第十章的規範將金融消費者金融保護的規範立法權從七個聯邦機構轉移至 CFPB[194]，因此原先 FRB 基於 ECOA 制定並實施的 Regulation B，於 2011 年 12 月時由 CFPB 於《美國聯邦法規》第 12 編第 1002 條以下重新制定 Regulation B 作為 ECOA 的實施細則。該規則第 1002.6(a)條即雷同 FRB 訂定的第 202.6(a)條規範，明訂禁止差別影響歧視，並將第 202.6(a)條註腳的效果測試概念寫於正文當中[195]。

第二款　FHA

---

[190] Miller, Fred H., *Note: A Perspective on the Regulation B 'Effects Test'*, THE BUSINESS LAWYER 62, no. 2, 560 (2007), http://www.jstor.org/stable/40688529 (last visited Jan 1, 2025).

[191] 整體而言，ECOA 與 Regulation B 皆體現了「經濟理性」（Economic Rationality）與杜絕「統計歧視」（Statistical Discrimination）的內涵。金融機構用以拒絕核貸的參考指標不應與其信用評分無關，亦不得僅因統計上被評價低信用者即拒絕該相關群體之貸款申請。

[192] Miller, Fred H., *supra* note 190, at 560-561.; *See* 40 Fed. Reg. 49298, 49301 (Oct. 22, 19)

[193] Winnie F. Taylor, *supra* note 189, at 580, 594, 629. https://brooklynworks.brooklaw.edu/jlp/vol26/iss2/3 (last visited Jan 2, 2025)

[194] 12 CFR §1002 (2011), https://www.federalregister.gov/documents/2011/12/21/2011-31714/equal-credit-opportunity-regulation-b (last visited Jan 3, 2025)

[195] *General rule concerning use of information.* Except as otherwise provided in the Act and this part, a creditor may consider any information obtained, so long as the information is not used to discriminate against an applicant on a prohibited basis. The legislative history of the Act indicates that the Congress intended an "effects test" concept, as outlined in the employment field by the Supreme Court in the cases of *Griggs* v. *Duke Power Co.*, 401 U.S. 424 (1971), and *Albemarle Paper Co.* v. *Moody*, 422 U.S. 405 (1975), to be applicable to a creditor's determination of creditworthiness.

doi:10.6342/NTU202500326

美國法下針對涉及房屋貸款之歧視主要由 FHA 規範。FHA 於 1968 年頒布，訂於《美國法典》第 42 編第 3601 條至第 3619 條，保障人民在租屋、購屋[196]、房屋抵押貸款、尋求住房援助等有關房地產交易與經紀服務時應免受歧視[197]，並明文列舉受歧視保護的群體為種族、膚色、宗教、性別、國籍、婚姻狀況或身心障礙等情形[198]。

美國住房和城市發展部（U.S. Department of Housing and Urban Development, HUD）是 Lyndon B. Johnson 總統根據住宅與都市發展部法案（Department of Housing and Urban Development Act, 42 U.S.C. 3532-3537）成立的內閣部門，其主要的任務是為了滿足美國的住房需求而負責國家政策和計劃，改善、發展住家社區，並執行和推動公平住房法[199]。2013 年 2 月 15 日，HUD 發布了關於 FHA 歧視性影響標準的實施規則「公平住房法案歧視性影響標準實施」（又稱 2013 規則）[200]，將長期以來的法院和機關見解明文化，確立了在 FHA 下也規範了差別影響歧視，此規則也藉由過往 FHA、HUD 的判例及民權法第七章的內容，制定了三階段的原告責任負擔移轉架構[201]，在這個架構中，第一，原告須初步證明被質疑的爭議行為已經或可預見地會導致歧視性影響；第二，若原告完成此初步證明，則被告必須證明該被質疑的行為係有助於達成其追求的合法利益，且利益與歧視無涉；第三，若被告在第二步驟中達成其舉證責任，原告仍可以進一步證明存有較少歧視性影響的方案，倘若該方案同樣能實現追求合法且無歧視的利益，原告仍可取得勝訴。

儘管 FHA 的規範係針對有擔保品的房屋貸款而與本文主要論述的信用貸款情形有間，然而在面對差別影響歧視的信用貸款歧視時，如前揭 HUD 有提出具體的原告責任移轉三步驟，在這個方案中縱使被告對歧視行為可以主張合理商業利益的抗辯，原告

196 42 U.S. Code § 3604
197 42 U.S. Code § 3605, §3606
198 *Id.* at 201.
199 *See* 42 U.S.C. 3608(a) and 42 U.S.C. 3614a.
200 78 FR 11460., “Implementation of the Fair Housing Act's Discriminatory Effects Standard” (“the 2013 Rule”).
201 78 FR 11460, 11482; *see, e.g., Inclusive Cmtys. Project, Inc.,* 576 U.S. at 527 (overviewing the 2013 Rule's burden shifting framework).

doi:10.6342/NTU202500326

可以進一步要求透過證明被告仍可以找到較小的歧視方案以追求合法利益以獲得有利的判決結果，本文認為縱使對於金融機構而言，有無擔保品直接影響貸款方案與決策結果的考量，然而 HUD 提出關於 FHA 原告與被告間的責任移轉架構，無論是對於房屋貸款或是信用貸款，是有助於被歧視者避免歧視者以追求合理商業利益的目的而規避歧視行為的責任，故在此考量下，應得以在信用貸款中考量援用 FHA 的相關實施規範，以尋求更好解決歧視的方案。

第二項　歧視認定之舉證責任、動機與因果關係

在 ECOA 和 FHA 中，關於歧視之動機與因果關係應如何解釋，有論者指出其與上述美國《1964 年民權法案》（Civil Rights Act of 1964）第七編（Title VII）的關係。民權法案是規範美國民權和勞動法上的指標性法案，其禁止在學校、工作場域及公共空間中的不平等對待。其中，Title VII 第 2000e-2(a)(1)條明定[202]：「禁止僱主在僱用、解雇、薪酬、晉升和其他僱傭方面，基於種族、膚色、宗教、性別或國籍等受保護特徵進行歧視。」該法是美國反歧視勞動法中範圍最廣且影響最深遠的，也因而設立一獨立性的平等就業機會委員會（Equal Employment Opportunity Commission, EEOC）以執行法定事宜[203]。目前多數見解認為民權法第七編禁止的歧視，同時包含了差別對待歧視與差別影響歧視[204]。

ECOA、FHA 和 Title VII 的規範模式大致相同，皆明列出規範所欲保護的特徵，但在實務上更常面臨的問題是應如何證明行為人具有歧視的動機及與歧視結果有因果關係。特別是針對訴訟程序中應由何方就何事實負有舉證責任，根據差別對待歧視或差別影響歧視而有所不同，以下針對美國法之重要判決所提出的舉證責任分配與模式討論。

---

[202] 42 U.S. Code § 2000e–2 (a)(1) r

[203] 焦興鎧（1990），〈美國勞工法上就業歧視之最新發展〉，《美國研究》，20 期，頁 59。

[204] 黃昭元，前揭註 167。；焦興鎧（1995），〈勞工法與勞工權利之保障〉，《美國勞工法論文集（一）》，頁 329、344-345，頁 344-45、353-54、363-64。

doi:10.6342/NTU202500326

第一款　差別對待歧視的舉證責任分配

1973 年美國聯邦最高法院麥道訴格林案[205]為飛機製造公司 McDonnell Douglas Corp. 資遣非裔美國人 Green 的案件。根據判決見解，若原告希望成立有立案可能「證據確鑿」（prima facie）的差別對待歧視案件，原告必須先負起提出以下證據的責任：(1)原告為 Title VII 下受保護之對象（protected group）；(2)原告之條件與資格符合僱主之要求；(3)儘管原告資格符合僱主的要求，仍被拒絕僱用；(4)原告被拒絕之後，該職位仍處於空缺狀態，且僱主繼續以同樣資格要求進行招募[206]。在原告提出之證據足以確立時，舉證責任則轉向被告，被告須提出正當理由抗辯其行為係為合法，而非基於歧視的理由[207]。

第二款　差別影響歧視的舉證責任分配

差別影響歧視的確立最早可見於 1971 年的 Griggs 案[208]，該案係以民權法第七編的 Section 703(a)(2)為依據[209]，以法律解釋的方式認定差別影響歧視也應是其所禁止的歧視類型[210]，從而確立差別影響歧視之正式法律依據[211]。依該判決之見解，除非行為人係基於業務上之需求（business necessity），否則以其他中性因素為理由而將受保護群體排除在外的行為，仍將構成歧視，且原告不需證明行為人具有歧視意圖（intent）[212]。

差別影響歧視因 Griggs 案而有正式的法院依據，然而差別影響歧視的構成要件在與舉證責任標準仍尚未有明確標準，直至 1991 年，美國國會修訂民權法將差別影響歧視正式列入第七編所規範的歧視種類之一，並在其中訂定了舉證責任。依其規定，原

---

[205] McDonnell Douglas Corp. v. Green, 411 U.S. 792 (1973).
[206] 翻譯節錄自：憲法法庭，《美國聯邦最高法院憲法判決選譯》，頁 182。
[207] 參見 Texas Department of Community Affairs v. Burdine
[208] *Supra* note 164.
[209] 42 U.S.C. § 2000e-2(a)(2).
[210] *Supra* note 164.
[211] 黃昭元，前揭註 167，頁 28。
[212] *Supra* note 164.

doi:10.6342/NTU202500326

告若欲提起差別影響歧視的案件，須先證明：1. 行為人的行為（措施）、2. 原告為受保護群體之人、以及 3. 行為人的行為構成差別影響結果[213]。

在承認差別影響亦為構成歧視的類型之一後，針對其認定的標準為何，Griggs 案中的法院係以統計方法來認定是否成立差別影響。然而，法院應如何認定個案中的差別影響達到統計上顯著（statistically significant）的程度？且具有統計上顯著之差別影響是否當然就構成差別影響進而而構成歧視[214]？仍有解釋上的疑義。

為評估是否存在可能的差別影響結果，在美國法常用的標準是由美國平等就業機會委員會（EEOC）在 1978 年發布的《統一員工選拔程序準則》中提出的「五分之四規則」（four-fifths rule）。據此，如果某一特定族裔或性別群體在選拔過程中的通過率低於最高通過率群體的 80%，則可能存在差別影響，需要進一步審查是否有正當理由，此規則為識別潛在的差別待遇之量化參考標準。

第四節　小結

本章從公平信貸的規範目的開始，首先探討契約自由與公平信貸間的價值取捨，進而討論公平、平等與正義間的差異，而以規範介入信貸中，是為了追求正義。接著從傳統以風險為考量的信貸審核中，觀察金融機構做成信貸決策的標準，以及常見的信用評估方法，觀察這些方法可能如何產生公平性的疑慮。

在規範面向，本章從立法模式開始，歸納出反歧視規範可分為封閉型、開放型及混合型三種模式，為了達到規範的明確性且同時仍保有解釋的開放空間，應採取混合型較為妥適。接著從規範中討論歧視的態樣，如美國法分為差別對待歧視及差別影響歧視，歐盟法分為直接歧視及間接歧視，梳理其定義與內涵，以及歧視的例外正當化事由。

---

[213] Civil Rights Act of 1991, 42 U.S.C. § 1981 (1991).
[214] 黃昭元，前揭註 167，頁 28。

doi:10.6342/NTU202500326

最後從比較法下的公平信貸規範，聚焦於美國的 ECOA、Regulation B 及 FHA 等規範，梳理這些規範中對於差別影響歧視的應用，以及責任移轉架構，並觀察過往實務上如何要求舉證責任分配，乃至認定具有歧視的具體標準。

doi:10.6342/NTU202500326

第四章 人工智慧信貸歧視之解決之道

在前一章中，透過比較法下美國公平信貸相關規範梳理其規範目的、反歧視規範的立法模式、歧視的類型以及允許歧視的正當化事由。本章將繼續以美國公平信貸規範為主，從人工智慧模型應用在信貸服務後，對現行的公平信貸規範可能會造成什麼樣的挑戰，並探討在法規範的框架下可以如何應對以避免人工智慧信貸歧視的發生。

第一節　人工智慧信貸的監管方式

為消除信貸中的歧視問題，若干國家已有針對公平信貸的法律規範，然而人工智慧信貸的歧視與傳統信貸歧視的癥結點不同，最顯著的兩個特徵為「不可解釋性」與「不可預測性」[215]。所謂不可解釋性意指，人工智慧系統的歧視結果的產生原因，是來自於人工智慧系統中的機器學習在特徵間尋找相關性而導致的結果，不同於人工審核的行為與結果間具有因果關係。人工智慧的不可解釋性主要係因其預測結果可能與開發該人工智慧模型的金融機構之行為或政策間沒有可解釋的關聯，甚至沒有因果關係，因人工智慧的學習能力，其可能會使用到金融機構的指令中所未預見的變數[216]。此外，複雜的模型（如神經網路）通常被認為難以解釋，儘管能夠得知模型的結構、資料輸入與權重比例，其行為也無法被解釋，使得使用者與開發者難以理解這些模型的決策邏輯[217]；所謂不可預測性，係指人工智慧系統的訓練目的是為了追求低風險且高利潤的貸款結果，此導致人工智慧系統產生歧視性結果的因素並非完全來自現行法的受保護特徵，任何影響風險和利潤結果的因素都可能成為演算法判斷的特徵之一。具體而言，許多人工智慧模型擁有千百萬個參數，使得其判斷行為難以預測，即使對模型的結構、資料輸入和權重有一定的了解，仍然無法準確預測模型的具體決策[218]。且人工智慧模型能夠根據新的資料輸入自我調整其決策模式，這種動態學習過程使得

[215] Wu, *supra* note 4, at 70.
[216] *Id.* at 4.
[217] *Id.* at 67.
[218] *Id.* at 60.

doi:10.6342/NTU202500326

其行為變得更加不可預測，也意味著模型在面對不同的輸入資料時，可能會產生意料外的結果[219]。人工智慧系統的運算過程往往不透明，使用者無法清楚了解模型是如何得出某一決策的，其也加劇了不可預測性[220]。

基於現行的反歧視法制的制定背景，係為了防止受保護特徵群體受不公平的對待的初衷，而這樣的受保護特徵在面對人工智慧信貸歧視時，可能發生不足以防範新型態歧視的情形[221]。又如上述人工智慧具有不可解釋性，故人工智慧模型的決策過成往往難以追溯，也因此受歧視者難以說明歧視結果與金融機構之行為間的因果關係[222]，亦難以舉證證明[223]。此外，金融機構所使用的人工智慧模型往往被視為商業機密，也因此難以被監管機構審查其是否帶有歧視性，因此也可能減弱了被規範的可能性[224]。

第一項　法規範

從法規範面向解決人工智慧系統的信貸歧視，可從兩個方向進行，一是訂定完整的金融反歧視規範，二是訂定人工智慧基本法。然而，應如何考慮規範的立法模式，是否得參酌傳統公平信用規範，將受保護特徵明定並保留空間之混合式立法模式，本文認為，在人工智慧系統與機器學習的使用情境中，這樣的規範體系將受到挑戰。演算法的決策過程是欠缺透明性且難以解釋的，以追求效率預測結果的運算過程中，可能出現訓練資料以中性特徵連結歧視的相關性，而產生帶有歧視性的決策結果。因此，透過列舉方式訂定反歧視法的受保護特徵，縱然具有表列該規範的核心價值，但在人工智慧的應用下，難以即時且全面地在法規中訂定明確、具有限定功能的受保護特徵。在人工智慧系統的信用貸款中，學者 Fuster 提出目前法規無法規範的情形為[225]：「金融機構雖無故意使用與保護特徵相關的中性特徵作為審核因素，然而演算法

[219] *Id.*
[220] *Id.* at 66.
[221] *Id.* at 1.
[222] *Id.* at 6.
[223] *Id.* at 6.
[224] *Id.* at 5.
[225] Fuster, *supra* note 108, at 54.

doi:10.6342/NTU202500326

透過歷史資料的分析與訓練，而找到與保護特徵具有相關性的負面因素（拒貸或金額低），卻因著演算法黑盒子的不透明性，使用者並未覺察而仍使用之。」儘管如此，將金融機構於金融服務中應受保護的群體特徵於規範中明訂，仍然有其必要性，因為在面對人工智慧信貸歧視時，我們希望防免的新型態歧視雖是以中性特徵為之，然而其實質上是與受保護特徵密切關聯，實際上仍然是對受保護群體的不公平對待。

首先，在金融反歧視法中須先訂定反歧視的規範原則，明定在金融服務當中不可針對受保護特徵群體有不公平的行為。其次，透過立法賦予金融機構強制揭露的義務，如《多德・弗蘭克華爾街改革與消費者保護法案》（Dodd–Frank Wall Street Reform and Consumer Protection Act）第 1033 條，要求金融機構必須向其消費者揭露用於消費者金融服務和產品的資料，以建立資料共享和機密審查的定期揭露制度[226]。

為解決演算法的不公平問題，雖可能導致影響演算法預期的表現的代價，但在公正與效率間的價值權衡下[227]，仍應從立法層面要求公平性。基於金融反歧視規範已明訂的法定受保護特徵，人工智慧基本法的規範應著重在透明的要求、問責制度與事後監督機制上，例如：針對人工智慧系統的說明文件、風險評估報告、安全性報告、責任歸屬架構.....等。

第二項　技術監管

除了開發者方監管，確保其開發的過程及方法符合規範，透過立法確定不公平、歧視的定義及受保護規範，從演算法的設計中預防排除。但此做法將大幅限制人工智慧系統的發展，因此可參考歐盟對於人工智慧系統的分類，根據產生的風險及潛在的風險分類，因應不同風險等級而有不同的監管要求。

[226] Wu, *supra* note 4, at 124.
[227] Odinet, *supra* note 23, at 9.

doi:10.6342/NTU202500326

根據目前的反歧視法規範的架構，若人工智慧系統使用受保護特徵作為評估標準時，此種直接歧視應直接被排除。然而學者 Andreeva 認為，在信用評分系統中過度限制變數的使用，可能影響其預測的準確性而對公平性有所挑戰[228]，而造成受保護群體被排除在使用變數外卻造成不利影響的原因，可能係因某些受保護群體面臨經濟上困境的風險較高，而有較高的違約可能性。

本文認為，在此論述的脈絡下，反歧視的目的是為了實現個人在信貸中的公平待遇，在直接對受保護特徵禁止的情形下，其他導致信用評分降低的變數權重反而升高，造成獲得係信貸機會減少的結果而加劇實質上的不公平。因此，確保其不會對受保護群體有不利影響，平衡信用評分系統的預測能力和效率也是反歧視規範需考量的面向[229]。

限制人工智慧系統的輸入資料，如要求排除受保護特徵，但演算法仍能由其他資料推斷出受保護特徵；若排除受保護特徵的替代指標，除了難以定義替代指標外，在變數相關性的推斷下亦可能連結至受保護特徵，以上從事前角度的規範多有缺陷而無法有效防免。因此，要求系統的透明性義務揭露，將其運算資料從結果面向觀察是否有公平性疑慮，並將其演算法模型與運算資料建立資料庫以持續監管，可能得以較全面地防範替代指標疑慮。

第三項　可解釋人工智慧與人力介入

第一款　可解釋人工智慧的原理

[228] Andreeva, Ansell, and Crook, *supra* note 81, at 22.
[229] *Id.* at 30.

doi:10.6342/NTU202500326

針對如何解決人工智慧歧視的問題，OECD 提出可從幾個面向著手[230]，其中包含加強人工智慧模式的可解釋性[231]。具體而言，人類設定人工智慧系統所追求的目標後，技術人員可在機器學習的訓練過程選擇輸入給系統學習的資料，並透過可解釋性的要求更加理解演算的模型、預期影響和結果，此將有助於建立有效的監管規範並防範潛在風險[232]。

針對可能發生的演算法風險，目前的應對方向大致可分為兩種：（1）提高演算法透明度；（2）對演算法的設計進行規範及監管[233]。以下介紹之：

第一目　提高演算法透明度

面對人工智慧偏見的問題，科技治理角度實可以「可解釋人工智慧」（Explainable Artificial Intelligence, XAI）解決之。可解釋人工智慧為解釋複雜模型決策的工具，目的是促進人類對人工智慧決策過程的理解，同時確保人類對該系統的信任，維持安全性及公平性[234]。此類「打開黑盒子」並解釋人工智慧決策過程的方法，縱使無法完全理解人工智慧的內涵，但也可藉此識別出風險[235]。學者 Keri 及 Joseph 也認為，現代人工智慧系統的「黑盒子」本質上是可以被解釋的，而可解釋的通常係以相對簡單的模型為主，如線性模型和遵循規則的模式[236]。

根據可解釋性的目的和需求而生的不同情境以及演算法的複雜度，可解釋人工智慧也發展出不同態樣，主要可分為外生（Exogenous Approach）及分解

[230] OECD (2022), OECD FRAMEWORK FOR THE CLASSIFICATION OF AI SYSTEMS, OECD DIGITAL ECONOMY PAPERS, https://doi.org/10.1787/cb6d9eca-en. (last visited Jan 1, 2025)
[231] Bloomberg, *Artificial Intelligence Has Some Explaining to Do* (Dec 12, 2018), https://www.bloomberg.com/news/articles/2018-12-12/artificial-intelligence-has-some-explaining-to-do (last visited Nov 2, 2024).
[232] *Id.*
[233] Keri Grieman & Joseph Early, *A Risk-Based Approach to AI Regulation: System Categorisation and Explainable AI Practices*, 61, 67, 69, 77, 81, 87 (2023), https://www.jearly.co.uk/papers/KG_XaiRegulation_SCRIPTed.pdf (last visited Jan 2, 2025)
[234] *Id.* at 59.
[235] Bar-Gill, Sunstein, and Talgam-Cohen, *supra* note 52, at 39.
[236] Grieman and Early, *supra* note 233, at 62.

doi:10.6342/NTU202500326

（Decompositional Approach）兩種方式[237]。外生方法指的是在演算法生成決策後，透過其他技術工具解釋決策，觀察並分析資料輸入和結果輸出之間的關係，進而解釋決策[238]；分解方法則是研究演算法內部的運作機制，對決策過程拆解分析，揭示得出決策的邏輯[239]。

監管機關可要求企業解釋並揭露其演算法資訊，為此監管機構需制定相關要求企業揭露的規範及程序，可能是揭露資訊或演算法代碼。然而，不透明的機器學習演算法日益普及，監管的實施將具有一定挑戰性[240]，且此舉可能有侵犯企業營業秘密之虞，且隨著資訊技術日益進步，強行揭露與解釋也可能出現無法解釋演算法之窘境。

第二目　介入演算法設計

相對於揭露人工智慧黑盒子並加以解釋，從人工智慧程式碼中加入反歧視的約束[241]，要求人工智慧設計方將該反歧視協議加入人工智慧系統之代碼中，也是可行的作法。具體作法可將法規範代碼化後輸入至人工智慧中，可以直接避免人工智慧對法定受保護特徵的歧視。

然而，縱使已將受保護特徵規範輸入至演算法中，仍無法有效防免歧視的發生，因為縱使不直接使用與受保護特徵相關的變數，仍可能透過其他變數（如教育程度、地理位置等）反映出與受保護特徵相關的歧視。

第二款　可解釋人工智慧之類型

可解釋人工智慧依據其適用的模型內部解釋性的複雜程度不同，可分為白盒模型及黑盒模型，以下就常見的可解釋人工智慧所適用的標的，分類敘述之。

---

[237] Ashley Deeks, *The Judicial Demand for Explainable Artificial Intelligence*, COLUMBIA LAW REV. 119, 1835 (2019).
[238] *Id.* at 1835.
[239] *Id.* at 1841.
[240] Bar-Gill, Sunstein, and Talgam-Cohen, *supra* note 52, at 4.
[241] Bar-Gill, Sunstein, and Talgam-Cohen, *supra* note 52, at 5.

doi:10.6342/NTU202500326

第一目　白盒模型

白盒模型（White-box Models）此類模型本身即具有可解釋性，結構相對易於理解且透明，例如：

一、線性迴歸（Linear Regression）：是一種用於尋找變數之間關係的統計技術，在機器學習環境中，線性迴歸可以尋找特徵和標籤之間的關係[242]。

二、決策樹（Decision Trees）：屬於監督式學習的一種，可用於分類和迴歸分析，其具有分層的樹狀結構，由根節點、分支、內部節點和葉節點組成[243]。

三、羅吉斯迴歸（Logistic Regression）：羅吉斯迴歸是一種監督式機器學習演算法，透過預測結果、事件或觀察的機率來完成二元分類任務，此模型提供僅限於兩種可能結果的二元或二分結果，例如是/否、0/1 或真/假，用以分析一個或多個自變數之間的關係，並將資料分類為離散類別[244]。羅吉斯迴歸廣泛用於預測建模，是金融機構最常用來信用評估的統計模型之一。

第二目　黑盒模型

黑盒模型（Black-box Models）此類模型的內部結構複雜且難以理解，無法估計第每個特徵對預測結果的重要性，也相對難以理解不同特徵間的相互關係，例如深度學習即屬於此類。針對黑盒模型常見可用的可解釋人工智慧解釋模型如下：

---

[242]Machine Learning, *Linear regression*, https://developers.google.com/machine-learning/crash-course/linear-regression (last visited Jan 1, 2025).

[243]JMP, *Introduction to machine learning: easy-to-understand interpretation of decision trees* (May 20, 2024), https://community.jmp.com/t5/JMP-Blog/%E6%A9%9F%E5%99%A8%E5%AD%B8%E7%BF%92%E5%85%A5%E9%96%80-%E9%80%9A%E4%BF%97%E6%98%93%E6%87%82%E8%A7%A3%E8%AE%80%E6%B1%BA%E7%AD%96%E6%A8%B9/ba-p/753561 (last visited Jan 5, 2025).

[244] SpiceWorks, *What Is Logistic Regression? Equation, Assumptions, Types, and Best Practices* (April 8, 2022), https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-logistic-regression/ (last visited Jan 1, 2025).

doi:10.6342/NTU202500326

一、個案代理模型（Local Surrogate）：其基本原理為透過可解釋模型以解釋個別預測結果，而非一次性解釋整個黑盒模型。代理模型又稱解釋器，其先在局部區域找到主要變數，再觀察各局部區域的主要變數是否有共通變數，進而找到整體模型的主要變數，可在資料集的局部區域生成簡單的可解釋模型來解釋黑盒模型採用的模型（例如上述提及的線性迴歸、決策樹等），LIME（Local Interpretable Model-agnostic Explanations）即是個案代理原理的模型[245]。

二、SHAP（SHapley Additive exPlanations）：Shapley values 係基於合作賽局理論（cooperative game theory）而生，用於計算賽局中的公平支出，也就是每個玩家對總支出的貢獻[246]。而 SHAP 則可用來估計 Shapley values，計算每個特徵對最終預測結果的貢獻值[247]。

三、反事實解釋（Counterfactual Explanations）：反事實解釋可用於解釋單一實例的預測，例如「事件」是實例的預測結果，「原因」是輸入到模型並「引起」某種預測的該實例的特定特徵值，即尋找會造成預測結果改變的最小變動變數群，進而鎖定模型當中主要使用的變數[248]。

第三目　其他類型

[245] Github, *Local Surrogate (LIME)*, https://christophm.github.io/interpretable-ml-book/lime.html. (last visited Jan 2, 2025).
[246] Petter Eilif De Lange et al., *Explainable AI for Credit Assessment in Banks*, 15 J. RISK FINANC. MANAG. 556, 4 (2022).
[247] 〈可解釋 AI (XAI) 系列 — SHAP〉，https://medium.com/ai-academy-taiwan/%E5%8F%AF%E8%A7%A3%E9%87%8B-ai-xai-%E7%B3%BB%E5%88%97-shap-2c600b4bdc9e（最後瀏覽日：05/08/2024）。
[248] Github, *Counterfactual Explanations*, https://christophm.github.io/interpretable-ml-book/counterfactual.html (last visited Jan 2, 2025).

doi:10.6342/NTU202500326

一、特徵重要性（Permutation Importance）：透過隨機置換個別特徵的數值，以評估該特徵對模型性能的影響，進而衡量特徵的重要性。其可用於各種機器學習模型，並不限於特定結構[249]。

二、師徒模式（Teacher-student）：亦稱為蒸餾（distillation），屬於可解釋人工智慧中的「代理模型」（Proxy Models）類型，透過訓練簡單且易於解釋的模型（學生模型），以模仿複雜且難以解釋的模型（教師模型）的行為，藉此輸出教師模型逼近預測結果值[250]。

第二節　公平信貸規範的新挑戰

第一項　民權法第七章、ECOA 與 Regulation B

如上一章所介紹，現行公平信貸規範強調的是因果關係（Causation）、合理可預見性以及緩解事前風險的制度[251]。但人工智慧系統中的機器學習有著高度的不透明性，無法完全解釋其決策的過程，導致適用現行的公平信貸法制規範人工智慧歧視的問題，可能存在實際上的困難。具體而言，雖然 ECOA 或 FHA 等規範中都沒有提及違法所需的意圖或因果關係的程度[252]，然而在美國最高法院的判決中可以發現，法院日益漸多援引侵權法的概念來解釋公平信貸法規[253]，在認定歧視上集中在隱含的意圖和因果關係，也因此原告負有證明行為人之行為與歧視結果間具有因果關係的舉證責任[254]。然而，這也導致了公平信貸法保護範圍限縮，因為侵權法的因果關係及舉證責

[249] Github, *Permutation Feature Importance*, https://christophm.github.io/interpretable-ml-book/feature-importance.html (last visited Jan 2, 2025).
[250] Medium, *Understanding the Essentials of Model Distillation in AI* (Jun 8, 2024), https://medium.com/stream-zero/understanding-the-essentials-of-model-distillation-in-ai-1e97403bee8a (last visited Jan 2, 2025).
[251] Wu, *supra* note 4, at 71.
[252] *See* 15 U.S.C. § 1691; *see also* 42 U.S.C. § 3601.
[253] Wu, *supra* note 4, at 103.
[254] *Price Waterhouse v. Hoplins*, 490 U.S. 228, 237(1989)

doi:10.6342/NTU202500326

任，在面對人工智慧歧視問題時，原告可能礙於人工智慧系統的不可預測與難以解釋，讓原告在證明人工智慧系統的決策導致歧視結果時，十分困難且成本高昂[255]。

如第三章所介紹，比較法上已有針對傳統信貸歧視的公平信貸規範，禁止金融機構在信貸服務中，使用法定受保護特徵對借款人在申請信貸的過程中為不公平的對待，而在使用人工智慧提供信貸服務時，縱使同樣要求人工智慧系統禁止使用公平信貸規範的法定受保護特徵，然而其可能為了達到貸款機構利益最大化、風險最小化的目標，使用了與法定受保護特徵的無關的變數，如前述在傳統信貸服務中已曾出現的劃紅線案例一樣的道理。但這個看似無關的變數的使用，實際上已經連結到規範上禁止的法定受保護特徵，卻可能因為人工智慧系統表面上所使用的變數並非規範明定禁止的法定受保護特徵，故縱使產生了歧視結果，也因為人工智慧使用的變數是表面上與法定受保護特徵無關變數，而難以被現行的公平信貸法規範。

須強調者為，在信貸流程中使用人工智慧系統並不一定會導致歧視的發生，也有可能因為機器學習比傳統人力判讀相比有更強的資料判讀能力，且使用客觀一致的貸款標準審核貸款人，反而緩和人力做成信貸決策時不同貸款專員彼此間的決策不一致性[256]。換言之，人工智慧信貸可能可以發現更多的個體資訊，其使用更為精準與客觀的風險預測技術，也可能有助於降低傳統人力核貸時對受保護特徵的依賴，故人工智慧信貸並不一定會惡化歧視問題[257]。

但傳統的公平信貸規範仍可能需要做出調整。傳統的公平信貸規範體系主要是以「輸入導向」的架構，也就是從資料輸入端直接要求金融機構排除受保護特徵的使用；然而傳統公平信貸規範以輸入為導向的立法模式套用至人工智慧信貸可能產生所

[255] Wu, *supra* note 4, at 109.; *See* Monette Davis, *Applying Twombly/Iqbal on Removal*, AM. BAR ASS'N. (Apr. 30, 2020), https://www.americanbar.org/groups/litigation/committees/pretrial-practicediscovery/practice/2020/applying-twombly-iqbal-on-removal/. (last visited Dec 12, 2024).
[256] Talia B. Gillis, *supra* note 83, at 1217.
[257] *Id.* at 1205-1210.

doi:10.6342/NTU202500326

謂的「輸入謬誤」（input fallacy）[258]，申言之，即使金融機構針對人工智慧的輸入資料已有嚴格的限制禁止其使用受保護特徵，然而縱使禁止人工智慧系統使用受保護特徵，人工智慧系統也會因演算法的運算而找到類似使用受保護特徵的結果，導致差別影響歧視。而在差別影響歧視中，因人工智慧之所以擇定這樣類似受保護特徵的目的，就是因為對於預測的結果有幫助，其產生決策結果的關鍵是基於相關性，且在差別影響歧視下，金融機構得以主張其係為了追求商業利益之正當性，而得以主張該人工智慧的使用係為正當目的，造成無法規範的差別影響歧視情形[259]。

針對上述困境，學者 Gillis 提出應以「結果導向」（outcome testing）的測試法，由監管機關透過該法測試人工智慧系統在信貸決策中是否有產生歧視的影響，透過實證的方式觀察其實際上決策的結果。具體運作上，首先需要擇定用於評估人工智慧系統結果指標，這些指標可以包括貸款核准率、違約率、不同群體之間獲得信貸的差異等，這些指標應能反映出人工智慧系統對不同受保護群體的影響[260]。同時，也需要準備包含借款人的資料集，且這些資料急需具有多樣化，而這些借款人的特徵應具有代表性（例如收入、信用歷史、地理位置等），資料集亦應包涵已知的貸款結果，以便進行比較和分析[261]。接著使用所選的機器學習演算法對目標資料集進行預測，針對每位借款人的信貸評分或貸款核准進行決策，此時應注意演算法的運算是透明的，並且能夠記錄每個預測的輸入特徵和結果[262]。最後再對機器學習的預測結果進行分析，特別是關注不同群體（例如不同種族、性別或社會經濟背景的借款人）之間的結果差異，可以透過統計方法來檢測是否存在顯著的差異，並評估這些差異是否符合公平信貸的標準[263]。

[258] *Id.* at 1180.
[259] *Id.* at 1185.
[260] *Id.* at 1245.
[261] *Id.* at 1181.
[262] *Id.* at 1247.
[263] *Id.* at 1185.

doi:10.6342/NTU202500326

透過上述三階段的測試，監管機關可發現金融機構使用的人工智慧系統是否對具有及不具有受保護特徵的群體間存在差別待遇，亦即擁有類似個人條件的借款人在應得到相同的平等對待，特別在於以人工智慧訂價與人為定價的差別影響程度相比較，進而觀察此人工智慧系統的定價規則是否增加或降低結果的不一致性[264]。結果導向的分析也是持續的過程，應定期檢查人工智慧系統的預測結果，並根據最新的資料和社會變化進行調整，以確保人工智慧系統不會隨著時間的推移而產生新的偏見或不公平[265]。透過上述結果導向的觀點，使監管機關在解決人工智慧信貸問題時，得更加關注在防止歧視影響的結果上，而避免許多傳統公平信貸法必須證明因果關聯性的窘境，降低許多實踐上的困難。

第二項　比較法上的人工智慧信貸歧視監管發展

如前揭章節所述，美國法已有制定完整且行之有年的公平信貸規範，亦有累積許多與信貸歧視相關的判決先例[266]，有鑒於信用貸款服務在美國人民消費生活中扮演的

---

[264] *Id.* at 1219.
[265] *Id.* at 1181.
[266] 然而目前似尚未見有法院針對人工智慧信貸歧視之公平性問題做出判決，較為接近的是 2023 年 12 月 7 日歐盟法院（CJEU）針對 SHUFA 使用自動決策系統預測信用風險所做出了一則重要判決（C.-634/21, OQ v. Land Hessen, with SCHUFA Holding AG.）。Schufa AG Holding（SHUFA）是德國一所私人信用紀錄評分機構，其蒐集德國境內居民進行的所有商業活動信用紀錄，並以此做出信用風險評分，做出評估個人信用度的結果，而這些評估結果將可能被企業或金融機構採計作為決定是否向某人發放貸款或與某人簽訂合約的風險評分。在案件事實中，SCHUFA 向金融機構提供預測個人未來行為的信用資訊，利用申貸者 OQ 的個人資料產生信用評分，並將此評分結果分享給德國銀行，最終銀行根據 SCHUFA 的評分結果拒絕了 OQ 的貸款申請，當 OQ 要求提供評定其信用分數之資訊時，SCHUFA 僅提供了分數及概述計算方法，並以商業機密為由拒絕揭露考量信用風險的具體因素 。歐盟法院在判決中表示 SHUFA 使用自動決策系統預測信用風險並分享給金融機構作為信用貸款服務評估的資料時，應根據 GDPR 第 22 條規定將相關資料揭露給相對人，其認為當 SCHUFA 的客戶僅根據評分值做出決策時，如銀行根據棋其分結果做出信貸決策，特別是當此評分結果影響第三方與個人建立或終止合約關係的決定時，該信用評分即構成自動決策。儘管 SCHUFA 主張其僅協助客戶決策而不是自己做決定，應不屬於 GDPR 第 22 條的範圍，但歐盟法院認為此舉已然構成 GDPR 第 22(1)條，該信用評分結果已相當於自動決策，具有法律效力或對資料主體產生重大影響。法院並進一步認為，消費者信貸申請中的分數不足將容易導致銀行拒絕申請，雖然 SCHUFA 本身並沒有做出拒絕貸款申請的決定，但在提供信用評分時，它對最終銀行的決策結果已發生了決定性作用。
從歐盟法院判決的立場觀察，法院似乎「寬認」了自動決策結果，而結合歐盟人工智慧法案以風險界定人工智慧系統的監督程度，將評估個人信用評分的人工智慧系統歸類為高風險人工智慧 ，法院在此判決認為任何利用人工智慧作為對個人產生重大影響的決策基礎的行為，都應遵守 GDPR 第 22(1)條，顯現了自動化決策系統中透明度的必要性，並強調了採用自動化決策系統的公司皆應揭露其使用自動化系統之相關資訊。

doi:10.6342/NTU202500326

重要性，以及美國長久以來的種族歧視問題，在人工智慧系統開始應用於信貸服務時，美國法奠基於其現有的公平信貸規範，能夠提出相對多元且可行的解決方案，也因此本文在探討人工智慧信貸歧視問題時，以美國法為比較法之主要參酌對象。

第一款　FTC

在美國法下，依據2010年Dodds-Frank法案，聯邦貿易委員會（FTC）具備ECOA和Regulation B的執法權限，從事相關的執法、研究和政策制定[267]。

2018年，美國汽車經銷商 Passport 被指控其在 2015 年至 2017 年間汽車融資中，對非裔人和拉丁裔的消費者收取較同類白人消費者更高的利率，且收取額外費用。在此案中，FTC 認為 Passport 上述有差別影響歧視結果的融資行為，已違反了《平等信貸機會法》[268][269]。在FTC的起訴書中，發現Passport公司在2017年8月至2020年8月間，對非裔消費者的平均利率加成為約28個基點（約291美元），而對拉丁裔消費者的平均利率加成為約26個基點（約235美元），非裔消費者獲得 Passport 允許的最大加成的比例比白人消費者高出約 47%，而相較拉丁裔消費者則高出約 38%。FTC 認為，融資的利率應是基於風險為相應的計算而決定，然而這些資料顯示非裔和拉丁裔消費者的利率高於非有色人種的白人消費者，並非基於其融資風險或是個人的其他信用特徵，已產生了統計上的顯著差異[270]，且 FTC 也認為這樣的這樣的作法並沒有為了

[267] *See* 15 U.S.C. § 1691c(c).
[268] Federal Trade Commission, *Federal Trade Commission Takes Action Against Passport Automotive Group for Illegally Charging Junk Fees and Discriminating Against Black and Latino Customers* (October 18, 2022), https://www.ftc.gov/news-events/news/press-releases/2022/10/federal-trade-commission-takes-action-against-passport-automotive-group-illegally-charging-junk-fees (last visited Nov 10, 2024).
[269] 有論者指出，透過FTC解決人工智慧信貸歧視的優勢是可以透過FTC的執法權力，以避免消費者需要證明行為人歧視的動機、或從演算法中找到歧視的原因，畢竟使用人工智慧系統的公司與受影響的消費者間可能存在權力不平等，消費者無法合理避免受到人工智慧系統的影響，有時甚至可能不知道其所受的不利影響係源於人工智慧系統； The Regulatory Review, *How the FTC Could Regulate Algorithmic Discrimination* (Jan 25, 2024), https://www.theregreview.org/2024/01/25/shaikh-how-the-ftc-could-regulate-algorithmic-discrimination/ (last visited Nov 10, 2024).
[270] *Federal Trade Commission, V. Passport Automotive Group, Inc.,* No. 8:2022cv02670 (US District Court for the District of Maryland)

doi:10.6342/NTU202500326

追求商業利益正當性，也不符合較小歧視替代方案，顯然已是基於種族、膚色及國籍等歧視性因素為不公平的結果。

第二款　CFPB

CFPB 亦有權針對任何違反聯邦公平放貸法（包括《平等信用機會法》）的實體提起公開執法行動，並可就公平放貸法相關事宜與其他聯邦及州級監管機構進行協調[271]。

根據 DFA 第 1033 條規定，消費性金融服務提供者必須向消費者提供其關於消費者使用金融產品獲服務的相關資訊。這可使消費者得以在信貸申請中，隨時要求提供金融機構使用人工智慧系統的具體資訊。CFPB 的 2023 年消費者保護報告進一步指出，根據 ECOA 與 Regulation B，貸款方對申請人採取不利行動時，需具體揭露拒絕授信的原因，且其理由必須與貸方實際考量或評分的因素相關[272]，此「不利行動通知」（adverse action notice, AAN）透過規定貸方主動解釋決策的方式[273]，以達成防止及識別信貸交易中歧視，這些通知也使消費者能夠了解貸方決策的具體原因，從而有機會改善其信用狀況或指出貸方不公平的錯誤決策。為實踐此不利通知，消費者金融保護局公佈了不利行動通知的範例表格[274]，內容包括貸方常用以考量拒絕授信的原因清單及開放欄位供貸方填寫未列出的原因。

在面對人工智慧信貸歧視時，如何判斷貸方是否已盡揭露原因的義務，此報告指出「特定性」（Specificity）的重要性，也就是若其揭露拒絕的原因與申請人沒有直接關連性時或不明確時，應具體詳細說明理由，而非僅列出拒絕的原因[275]。

---

[271] *Id.* at 94.
[272] 12 CFR Part 1002 (Supp. I), sec. 1002.9, para. 9(b)(2)-2.
[273] 15 U.S.C. 1691(d)(3); 12 CFR 1002.9(b)(2).
[274] 12 CFR Part 1002, (App. C), comment 3.
[275] 12 CFR Part 1002 (Supp. I), sec. 1002.9, para. 9(b)(2)-4.

doi:10.6342/NTU202500326

另一方面，《住房抵押貸款揭露法案》（The Home Mortgage Disclosure Act, HMDA）也在實現公平信貸中扮演著重要的角色。為解決住房房貸中的不平等問題，《住房抵押貸款揭露法案》要求金融機構對於貸款申請、授信等詳細資訊有報告義務，包括申請人的種族、性別、收入、貸款金額和貸款用途等，而這些資訊也成為監管機關監管的基礎，得以識別潛在的歧視風險，CFPB作為監管機關之一，透過此法案所揭露的資料，得以評估金融機構的貸款程序，在發現歧視現象時得以介入[276]。

在面對人工智慧信貸歧視時，《住房抵押貸款揭露法案》的資訊揭露解決了演算法不透明性所帶來的困境。透過其揭露的資訊，監管機關得以連結既有的公平信貸規範，以發現金融機構在信貸流程中的歧視行為，並透由監管機關的職權調查，可以對其警告或要求改正，亦可能公布調查結果以揭露其歧視行為。在嚴重的情況下監管機關甚至可將其轉介給司法機關，例如平等機會信用法第 704 條規定，負責執法的機構在有理由相信借貸方有系統性或貸款歧視時，必須將其移交給美國司法部（DOJ）處理[277]。

第三節　較小歧視替代方案與可解釋人工智慧的應用

第一項　較小歧視替代方案的內涵

2024 年 6 月，消費者報告（Consumer Reports）和美國消費者聯盟（Consumer Federation of America）致信給消費者金融保護局（CFPB），敦促該局發佈指引，明確規範金融機構有義務在消費者信貸授信和定價中減少使用人工智慧系統產生的歧視性不利影響[278]。其並建議 CFPB 應要求使用自動決策工具的公司符合《平等信貸機會

[276] CONSUMER FINANCIAL PROTECTION BUREAU, FAIR LENDING REPORT OF THE CONSUMER FINANCIAL PROTECTION BUREAU, 7,14 (JUN 26, 2024), https://files.consumerfinance.gov/f/documents/cfpb_fair-lending-report_fy-2023.pdf (last visited Jan 5, 2025)
[277] *Id.* at 34.; 15 U.S.C. § 1691e(g).
[278] Consumer Federation of America, *Consumer Groups Call on CFPB to Protect Consumers from Discriminatory Algorithms Used by Banks and Other Financial Institutions to Make Credit Decision* (June 26, 2024), https://consumerfed.org/press_release/consumer-groups-call-on-cfpb-to-protect-consumers-from-discriminatory-algorithms-used-by-banks-and-other-financial-institutions-to-make-credit-decisions/ (last visited Nov 10, 2024).

doi:10.6342/NTU202500326

法》等反歧視法規，主動減少差別影響歧視的發生[279]。該信還建議 CFPB 在以下方面提供清晰指導：例如明確指導貸款人如何在模型開發過程的每個階段主動減少差別影響、持續提供指導，明定搜尋較小歧視替代方案的方法，包括適當的技術、頻率和強度、提供適當的公平性測量指標和方法、明定哪些替代方案是可行的[280]。

CFPB 基此指出，使用人工智慧系統以進行信貸審核的公司應尋找「較小歧視替代方案（Less discriminatory alternatives, LDA）」，但在未有明確的指引發布前，其尚無法要求各金融機構達到此要求[281]。

較小歧視替代方案的法律基礎源於《民權法第七編》，其禁止雇主基於種族、顏色、宗教、性別或國籍的就業歧視，故雇主必須考慮其決策對不同群體的影響，若存在歧視影響，原告可以提出較小歧視替代方案要求以作為替代方案[282]。在 1991 年民權法修正案中，第七篇有了重大的修改，特別是與差別影響歧視的證明有關，修正案強調了雇主在面對歧視指控時，必須能夠證明其決策的有效性，且必須考慮較小歧視替代方案[283]。

美國最高法院亦在許多案件中承認較小歧視替代方案的應用，如 Griggs v. Duke Power Co.案中，最高法院認為當雇主訂定的政策對某些特定族群產生差別影響時，其同時應考慮是否有同樣能達成其合法商業利益且不具有歧視性的其他方案[284]；在 Ricci v. DeStefano 案中，法院亦強調雇主在面對潛在的歧視指控時，必須有充分的證據來支持其不公平行為，並應主動提出較小歧視替代方案，判決並支持較小歧視替代方案的合法性[285]，較小歧視替代方案也體現在上述的《公平住房法》三階段步驟中。

---

[279] *Id.*
[280] *Id.*
[281] *Id.*
[282] Allan G King & Alice H Wang, *Artificial Intelligence As A Less Discriminatory Alternative*, 33 PUBLIC POLICY, 376 (2023), https://scholarship.law.ufl.edu/jlpp/vol33/iss3/2/ (last visited Nov 10, 2024)
[283] *Id.* at 14.
[284] *Supra* note 164.
[285] *Id.*; *Ricci v. Destefano*, 530 F.3d 88 (2d Cir. 2008)

doi:10.6342/NTU202500326

在面對演算法歧視問題，2019 年 8 月美國住房與城市發展部（U.S. Department of Housing and Urban Development, HUD）提出了新的建議規則以評估因住房房貸引發的演算法歧視問題，該規則於於 2021 年 6 月生效，其提出應以五要素主張取代 HUD 先前的三步驟責任轉移推定架構，原告必須主張 1.「被質疑的政策或做法是恣意的，且非達成有效利益或合法目標（如實際商業利益、盈利、政策考量或法律要求）所必須的」；2.「被質疑的政策或做法與對受保護特徵群體所受的差別影響間存在緊密的因果關係」；3. 被質疑的政策或做法對受保護類別成員造成「不利影響」；4.「該政策或做法所引起的差異顯著」；以及 5. 原告所指控的歧視係直接由被質疑的政策或做法造成」，[286]根據此主張架構，被告無法以追求合法商業利益且無「較少歧視性替代方案」的主張來否決原告認為差別影響歧視的主張。

第二項　人工智慧信貸與較小歧視替代方案

較小歧視替代方案如套用至人工智慧信貸的具體實踐上，首先是金融機構需先評估現有的運作系統是否存在對某些受保護群體不利的差別影響，檢查其決策結果是否存在某些受保護特徵的顯著差異性[287]。當發現差別影響存在時，金融機構應在滿足其追求合法正當商業利益的條件下，提出較小歧視替代方案以減少對特定群體的不利影響。

針對金融機構如何找出較小歧視替代方案，有學者提出可以人工智慧分析大量資料幫助金融機構找尋潛在的較小歧視替代方案，透過人工智慧系統找到兼顧公平性及有效性的平衡[288]。當金融機構選擇了較小歧視替代方案並開始踐行後，仍然要持續監測結果，確保其能有效減少歧視性的影響，並且應不斷進行優化，持續確保公平性。

---

[286] U.S. DEPT. OF HOUS. & URB. DEV., HUD's Implementation of the Fair Housing Act's Disparate Impact Standard. *A Proposed Rule by the Housing and Urban Development Departmen*t, 84 FED. REG. 42854 (Aug. 19, 2019).
[287] King and Wang, *supra* note, 282 at 378.
[288] *Id.* at 26, 30.

doi:10.6342/NTU202500326

為促使人工智慧信貸應用較小歧視替代方案，監管機關首先應明確定義「較小歧視」的意涵，並制定公平性標準以評估其替代模型是否足以達成公平性的目標，例如以不利影響比例（adverse-impact ratios, AIRs）作為評估標準[289]，AIRs 是受保護群體的貸款核准率除以對照組的貸款核准率的比率（基於信用模型產生的分數）[290]，如果 AIRs 值低於特定的「實際顯著」閾值，可能就具有統計顯著性[291]，如技術人員依據 EEOC 的五分之四規則制定了小於 0.8 的 AIRs 閾值[292]，用以確定是否存在實際重大的不利影響。

再者，金融機構選擇較小歧視替代方案時，應考慮群體間的特徵差異性，具體而言，當金融機構使用人工智慧系統進行決策時，必須評估模型的表現是否因性別或其他群體特徵而有所不同，在進行歧視影響分析時，金融機構需要在使用前後持續進行評估，確保不會對某些群體造成不成比例的影響[293]。除了追求公平性之外，金融機構合理應追求的商業利益也不應被過度犧牲，故如何兼顧其中的平衡亦應為考量因素之一[294]。

第三項　可解釋人工智慧的應用實例

在人工智慧信貸系統中，差別影響歧視是目前難以透過受保護特徵的分類規範，然而本文認為可以使用可解釋人工智慧來識別其中的歧視。如前述的線性回歸、決策

---

[289] Office of the comptroller of the Currency*, Comptroller's Handbook: Fair Lending* (2023), https://www.occ.treas.gov/publications-and-resources/publications/comptrollers-handbook/files/fair-lending/index-fair-lending.html. (last visited Nov 10, 2024)
[290] Richard Pace, *Algorithmic Justice: What's Wrong With The Technologists' Credit Model Disparate Impact Framework*, PACE ANALYTICS CONSULTING LLC, (May 21, 2024), https://www.paceanalyticsllc.com/post/ai-credit-model-disparate-impact#viewer-jhqqa2223732 (last visited Oct 12, 2024)
[291] *Id.*
[292] U.S. Equal Employment Opportunity Commission, *Select Issues: Assessing Adverse Impact in Software, Algorithms, and Artificial Intelligence Used in Employment Selection Procedures Under Title VII of the Civil Rights Act of 1964*, https://www.eeoc.gov/laws/guidance/select-issues-assessing-adverse-impact-software-algorithms-and-artificial#_ednref21:~:text=What%20is%20the%20%E2%80%9Cfour%2Dfifths%20rule%E2%80%9D%3F (last visited Jan 2, 2025)
[293] Emily Black et al., *Less Discriminatory Algorithms*, 101 GEORGETOWN LAW JOURNAL, 96 (2023), https://www.law.georgetown.edu/georgetown-law-journal/wp-content/uploads/sites/26/2024/11/Black-et-al._LessDiscriminatory.pdf (last visited Nov 2, 2024)
[294] *Id.* at.4

doi:10.6342/NTU202500326

樹、羅吉斯回歸，可以透過此類模型分析人工智慧系統對某些特定群體（如不同性別或種族）的輸出結果，檢查是否存在系統性的偏差；而 SHAP 方法提供了局部解釋能力，可以透過分析單一樣本的預測，顯示特定特徵對結果的貢獻程度，可以用以幫助判斷模型是否過度依賴於敏感特徵，特徵重要性分析亦可以幫助確定模型是否過於依賴於潛在的敏感特徵；而反事實解釋則可以模擬如果改變這些特徵，結果是否會有顯著變化，例如若想知道借款人為何被拒絕貸款請求，可以將表述問題為反事實，將預測從拒絕變為核准的特徵（例如收入、信用卡數量、年齡等），進而觀察最小變化是什麼而找到主要的變數。

Upstart Holdings 即是人工智慧信貸使用較小歧視替代方案的實際案例之一。Upstart Holdings 是一個提供人工智慧借貸平台的金融科技公司，其透過大數據、非傳統申請者資料（如借款人的高等教育背景）及人工智慧來審核借款人，以評估信用風險決定是否核貸及設定貸款利率。由於其演算法可能存在種族歧視的風險，Upstart 同意讓監測機構評估其演算法[295]。監測機構對 Upstart 的模型測試後，發現該模型對非裔申請者的核貸率明顯低於非西班牙裔白人申請者，引起對該模型潛在歧視風險的擔憂。

在發現模型存在針對種族的差別影響後，監測機構開始尋找是否存在較小歧視的替代方案，其主要透過兩種主要程序來尋找替代方案[296]。首先，監測機構檢查了原模型輸入特徵的所有可能子集，以識別能夠減少歧視性影響的組合，此為特徵選擇（feature selection）。例如，如果原模型使用特徵 A、B 和 C，監測機構會測試僅使用 A 和 B、A 和 C、B 和 C 等組合[297]。另外，監測機構還進行了超參數調整（hyperparameter tuning），係指找到最佳的超參數值以優化模型架構或訓練過程。監測機構亦使用不確定性區間來評估替代模型的性能（uncertainty assessment），此方法提供了對模型預測

[295] Emily Black et al., *supra* note 293, at 113.
[296] *Id.* at 35.
[297] *Id.* at 35.

doi:10.6342/NTU202500326

結果的量化理解，有助於解釋模型的預測結果，並提供了對模型可信度。透過這些方法，監測機構能夠識別出一些較少歧視的替代模型，這些模型在保持預測準確性的同時，顯著減少了對特定群體的歧視性影響[298]，Upstart Holdings 的案例展現在信貸領域中主動尋找並實踐較小歧視替代方案的重要性

第四節　小結

在本章中，從規範傳統信貸的公平借貸規範出發，梳理如 ECOA、Regulation B 在人工智慧信貸系統中，是否仍得以適用？進而觀察美國監理機關如 FTC、CFPB 在面對人工智慧信貸歧視時的監管發展。

本文認為，從美國 FHA 參酌 HUD 的較小歧視替代方案，要求金融機構負有尋找較小歧視替代方案，可以作為解決傳統公平信貸規範無法防免人工智慧系統所帶來的新型態歧視。而在具體操作上，可以要求金融機構使用可解釋人工智慧，用以識別其人工智慧系統中是否造成歧視？特別是在差別影響歧視中，應識別出是否有使用到敏感的特徵。此外，監理機關亦可以使用可解釋人工智慧監管金融機構所使用的人工智慧系統，檢視其是否帶有歧視。

傳統公平信貸法規範行為不應使用受保護特徵的面向，然而在人工智慧系統中，從輸入資料的面向限制並無法防止模型尋找類似受保護特徵的其他特徵，而仍然複製了與使用受保護特徵相同的結果。因此本文認為，藉由較小歧視替代方案與可解釋人工智慧的使用，可以從結果面向避免人工智慧信貸歧視的發生，除了要求金融機構的自我檢驗外，也提供給監管機關從外部監督的工具。

[298] *Id.* at 36.

doi:10.6342/NTU202500326

第五章 我國法下之人工智慧與公平信貸規範

前一章從美國法對於人工智慧歧視的各種討論中，提出了以較小歧視方案及可解釋人工智慧作為人工智慧信貸歧視的解決方案。本章將從我國法的現況檢視，應如何將此方案妥為適用，並提出相關的芻議。

第一節　人工智慧信貸歧視的控管方向

從前揭章節的討論中可以發現，在金融機構開始應用人工智慧系統提供信貸服務後，人工智慧信貸歧視的案例日益顯見，國際上無論是政府機關或是學者間，皆不斷嘗試在現行的公平信貸規範下討論，試圖找出得以援用的現行法律，並提出可能面臨的限制，或提出不同於現行規範架構的監管路徑，以解決人工智慧系統應用在信貸中所產生的不公平結果。

人工智慧信貸歧視對監理帶來的挑戰是其經常使用替代特徵產生事實上的歧視結果，導致傳統僅禁止使用受保護特徵的反歧視法不當然能解決問題。為有效規範人工智慧信貸歧視，本文認為由美國公平信貸規範創建的差別影響歧視概念，或許是目前最能解決此種使用表面上中性特徵的「代理」歧視的主要路徑。

然而將差別影響歧視實際應用至人工智慧信貸時，仍存在許多困難，例如傳統公平信貸規範在認定是否構成歧視時，一方面須證明使用中性特徵的行為與歧視影響間的因果關係，另方面行為人可以其係追求合理正當利益為理由正當化其差別影響歧視，此正當理由抗辯即可能使反歧視規範無法發揮效用[299]。而礙於人工智慧系統的不透明性，在機器學習的輸入資料與決策結果間，通常難以或需耗費許多成本才得以找到資料與歧視結果的因果關係，而僅能找到輸入資料和結果間可能存在某些程度上的相關性。然而，如第四章所述，可解釋人工智慧提供了金融機構以及主管機關一個偵

[299] Frederik J. Zuiderveen Borgesius, *Strengthening Legal Protection against Discrimination by Algorithms and Artificial Intelligence*, 24 INT. J. HUM. RIGHTS 1572, 7 (2020).

doi:10.6342/NTU202500326

測人工智慧信貸是否有歧視行為的管道，可透過要求金融機構運用可解釋人工智慧，檢測其所使用的人工智慧信貸系統是否存在歧視，主管機關亦可以運用可解釋人工智慧去檢查金融機構所使用的人工智慧信貸系統是否存在歧視。

反觀我國現況，我國雖尚未出現人工智慧信貸歧視的典型案例，然而在金融機構應用人工智慧系統日亦漸長的趨勢下，仍有未雨綢繆討論的必要性。隨著各國的人工智慧規範、資料保護法制的發展，我國金管會如上述也已參考國際規範與比較法規範後公佈《金融業運用人工智慧（AI）指引》，是目前針對金融機構應用人工智慧工具相對完整的指引，銀行公會也公佈有《金融機構運用人工智慧技術作業規範》，要求公會會員於運用人工智慧時應遵循的規範，試圖防範金融服務中的人工智慧歧視。然而這些指引或規範距離上述控管人工智慧信貸的方向仍有不足之處。故本文將在以下的討論中，借鑒、整合目前比較法上的規範解釋，針對我國現行規範提出建議。

第二節　制訂信貸反歧視法之芻議

第一項　制定規範的必要性討論

如何針對人工智慧信貸歧視制訂有效的立法規範？仍有討論空間。特別是人工智慧信貸歧視主要來自其大量運用中性特徵導致歧視結果，但法律規範難以將所有中性特徵涵括為人工智慧信貸系統不可使用的特徵，導致難以界定使用的規範界線。此外在人工智慧金融正蓬勃發展時，對其應用設有過多的限制，也可能導致企業遵法成本提升，進而影響創新發展[300]。另一方面，使用人工智慧系統減低了對信用評估模型的依賴（如 FICO），也可降低信用隱形的問題[301]，故是否需要對人工智慧信貸歧視有過分的擔憂而祭出嚴格的監管？應可繼續觀察。

[300] 廖淑君，前揭註 53，頁 144。

[301] *See generally* William Adams, Liran Einav & Jonathan Levin, *Liquidity Constraints and Imperfect Information in Subprime Lending*, 99 AM. ECON. REV. 49-84 (2009); Liran Einav, Mark Jenkins & Jonathan Levin, *Contract Pricing in Consumer Credit Markets*, 80 ECONOMETRICA 1387-1432 (2012); Liran Einav, Mark Jenkins & Jonathan Levin, *The Impact of Credit Scoring on Consumer Lending*, 44 RAND J. ECON. 249-274 (2013).

doi:10.6342/NTU202500326

本文考量上述正反面向後，認為使用人工智慧系統之原意應是提高效率、為人民創造更多福祉，雖然立法限制會造成金融機構的成本，但人工智慧信貸既已出現新型態的歧視問題，已偏離使用人工智慧的原意，透過立法避免問題的發生應是可考慮的方法之一。然而，考量人工智慧系統的持續發展和實證上的歧視結果，本文也認為立法僅需針對反歧視原則與法定受保護特徵作定義性與原則性的規範即可，換言之仍應給予系統提供者適當的空間及行為自由，不宜過分限制。

本文在第二章第三節探討我國現行法下可能得以作為公平信貸的規範依據，目前我國關於金融機構應用人工智慧的規範，主要在《金融機構運用人工智慧技術作業規範》和《金融業運用人工智慧（AI）指引》中有較為完整的規範內容，可能得以此作為解決人工智慧信貸歧視問題的法源依據，本文在接下來的討論中將針對此二規範的內容，提出可增加或修正的建議。

第二項　規範層級及規範對象

在效力方面，《金融機構運用人工智慧技術作業規範》的定位屬於公會的自律規範，依據該法第 15 條：「金融機構應將本規範內容，納入內控內稽制度中，並定期辦理查核。」其係要求銀行業間自律遵守，並藉由公會組織體系以內部自我管制與約束，故本質上較為自律規範的要求，並不具有法律或命令的拘束力地位；而《金融業運用人工智慧（AI）指引》則在前言中開宗明義表示，該指引之性質為行政指導，不具有拘束力。相較前者自律規範尚有內控內稽制度的自我約束要求，後者的約束力更為薄弱，似乎表明此指引僅供參考，金融機構沒有遵守的義務。

然而，金融服務中的公平性影響金融消費者的權益甚大，人工智慧系統應用在金融服務中可能發生潛在的風險，導致公平性產生疑慮，已是不可忽略的隱憂，故本文認為應將金融機構運用人工智慧技術的規範及指引內容，且應以法令位階的形式規範，具有一定的拘束力，才能確實要求金融機構有遵守的義務，使機構重視此議題。

doi:10.6342/NTU202500326

再就規範對象而言，人工智慧在金融服務中得應用的面向眾多，信貸的信用評分、定價及授信僅是其中一種應用成果，本文雖係聚焦於人工智慧應用於信貸服務的情形，然而整體而言，我國如上述根本性地缺乏金融反歧視法的一般性規範，因此本文建議在設定立法的規範對象時，不應僅侷限於銀行業者與信貸服務，而應針對更廣泛的金融機構與金融服務為規範對象，以將更多可能的潛在議題囊括。

第三項　金融反歧視法的基本規範內容與法定受保護特徵

金融機構與消費者間的金融服務係基於私法上的契約關係，在未有金融反歧視規範訂定前，僅能透過《民法》第 72 條概括條款，以維護誠信原則、公序良俗等不確定法律概念將《憲法》第 7 條平等權的精神引入民事契約關係中，基於利益衡量的目的兼顧民事關係上的權利義務與基本權保障[302]。然而就立法論而言，為落實憲法上對基本權的保障，本文如上述認為應於金融領域訂定一般性的反歧視規範，在契約自由與人權保障間取得平衡，以在私法自治原則下透過反歧視的保護實現實質上的自由[303]。

我國如上述並無一般性的公平信貸規範，因此對於信貸歧視欠缺基本規範，也無針對歧視的態樣有如美國《平等機會信用法》之差別對待歧視及差別影響歧視之規範，縱使《金融消費者保護法》第 7 條第 1 項中有公平合理、平等互惠及誠信原則的要求，仍無法因此推論出明確的受保護特徵。此為我國法欲開展人工智慧信貸歧視規範首先面臨的法制不足。

誠然如上述，人工智慧信貸歧視的挑戰之處在於人工智慧可能使用中性替代特徵達到歧視結果，故公平信貸法律需要重新構思歧視的概念。但在此之前，本文認為我國法仍應先訂定明確的受保護特徵，明定在金融服務中應禁止歧視的規範，作為金融服務應維持公平性的具體標準。

---

[302] 葉啟洲，前揭註 120，頁 41。
[303] 葉啟洲，前揭註 120，頁 42。

doi:10.6342/NTU202500326

受保護特徵的選擇是價值衡量和立法政策的問題[304]，從目前的反歧視法分析受保護特徵的擇定，如前述可歸納為「不變性、相關性和任意性、歷史壓迫和結構劣勢、社會特性」四種分類方式。考量法規範的明確性與可預測性，本文認為應參考兩公約與各國反歧視法規範，針對具有結構性劣勢且涉及個人追求自由生活所面臨的限制，作為我國法下法定受保護特徵的制訂方向，例如：性別、種族、膚色、民族、遺傳特徵、語言、宗教、政治傾向、殘疾、年齡或性取向等分類。考量保護的周延與具體性，本文亦建議我國應可採用「混合式」立法結構，透過例示將目前多數反歧視規範認定的受保護特徵列入，但仍保留未來納管「與之相似的」其他特徵的空間。

《金融機構運用人工智慧技術作業規範》第六條指出：「金融機構運用人工智慧時，在演算法設計、開發、資料蒐集、訓練資料選擇、處理、模型建置/生成/優化，及後續應用於金融服務過程中，應採取措施以符合金融服務業公平待客原則。於處理人工智慧不公正或偏見問題時，以下資料參數得評估是否納入演算法判斷，如：姓名、法律無限制或禁止之年齡、所有生理特徵（包含且不限於身高、體重、性別、膚色、髮量、肢體障礙等）、所有非涉及心神喪失致無法自主理解該金融商品判斷能力之疾病，並應就資安、法遵及風控等層面評估風險，依內部程序辦理。」此外《金融業運用人工智慧（AI）指引》第二章第二、（一）、1 點指出：「如無合理理由，運用 AI 系統所產生之決策則不應對特定群體有系統性之不利差別待遇（例如不得以特定宗教、種族、性別、身心障礙、性傾向、居住所、政治傾向、年齡、國籍或族群等因素，對借款人提供不合理的貸款條件）。」兩者均有例示一定的受保護特徵，但兩者的內容又不盡相同。

[304] 葉啟洲，前揭註 120，頁 45。

doi:10.6342/NTU202500326

本文認為，後者所規範的受保護特徵相較前者較為完整且合理，亦是比較法上常見傳統的受保護特徵，故本文建議得維持此規範，並參酌「混合式」立法結構，仍保留未來納入其他相似特徵的空間，以因應人工智慧歧視的不確定性。

第四項　正當理由例外與較小歧視替代方案

《金融業運用人工智慧（AI）指引》第二章提出「重視公平性及以人為本的價值觀」的核心原則二，指出金融機構在使用人工智慧系統時，應盡可能避免演算法偏見所帶來的不公平，並定義公平性為「金融機構運用 AI 系統產生之決策，不應對特定群體造成歧視之結果，亦即決策需有合理性、準確性及儘可能避免歧視。」，此外其針對決策合理性進一步規定「如利用個人屬性做為 AI 模型決策之因素之一，應有合理理由」。指引將有合理理由作為人工智慧金融歧視的例外正當理由，與國際上發展出的規範內容類似，值得肯定；但指引對此「合理理由」的認定並無具體說明，為其不足之處。

如前揭章節所述，美國公平信貸法對於使用受保護特徵而產生歧視、不利影響時，只要可以證明其係追求合法正當的利益，即可認有正當理由而豁免歧視責任。然而，《公平住房法》進一步提出較小歧視替代方案概念，強調儘管行為人得以舉證說明歧視行為係為了追求合法的商業利益，但受歧視者仍可透過證明行為人得選擇較小歧視的替代方案，而使行為人仍須為其歧視行為負責。而在近期的討論中亦有論者認為，這樣的較小歧視替代方案可透過美國 CFPB 以監管機關的立場，要求金融機構若欲使用受保護特徵而造成不利影響時，除了證明其係為追求合理的商業利益，如判斷借款人的信用風險，亦應同時證明其無其他可使用的較小歧視替代方案[305]。

本文認為正當合理理由例外可能過於廣泛，造成金融機構只要聲稱其使用個人特徵作為決策資料係為了追求其商業利益，通常都可以主張正當合理理由例外。故本文

[305] Consumer Federation of America, *supra* note 278.

doi:10.6342/NTU202500326

建議在《金融業運用人工智慧（AI）指引》的規範內容中，應可參酌美國法下較小歧視替代方案的概念，除了合理理由外，亦要求金融機構有義務說明其已盡力而無法找到其他較小歧視替代方案以達成其追求的目標，方屬盡其說理義務。就此，指引第二章第三、（四）、2 點「金融機構宜辨識運用 AI 系統與受系統性不利差別待遇之特定群體間是否具有關聯性，如是，金融機構宜採行降低該特定群體受影響之方式」的規定，似可作為課予金融機構考量較小歧視替代方案責任的基礎。

第五項　主管機關的調查監督

《金融業運用人工智慧（AI）指引》在第二章第三點說明公平性的落實方法時，分別從「系統規劃及設計」階段、「資料蒐集及輸入」階段、「模型建立及驗證」階段及「系統部署及監控」階段提出防範公平性問題的具體操作指引，在監管方面係提出由金融機構宜定期檢視與分析人工智慧系統產出之結果是否存在歧視，若發現歧視問題，應及時進行調整改進。此規定要求金融機構應定期檢視其使用的人工智慧系統的公平性，值得肯定。

值得補充者為，如前揭章節所述，美國 CFPB 及 FTC 作為監管機關，在金融機構運用人工智慧系統做成帶有歧視的決策時，得介入要求金融機構揭露其決策之具體內容，並說明其具有合理正當理由及沒有其他較小歧視方案得選擇的說明。我國目前《金融機構運用人工智慧技術作業規範》及《金融業運用人工智慧（AI）指引》並非具有強制力的法律規範，可能缺乏賦予監理機關得介入要求金融機構的正當基礎，故本文建議可參酌美國法規範，賦予金管會得在具體事件中介入調查的權力，並在確認造成歧視結果時得要求金融機構停止不當行為。

doi:10.6342/NTU202500326

第六章 結論

平等原則是現代生活中普世原則，也是保障人民追求社會生活的重要基礎。隨著技術的快速發展，人工智慧系統近年成為各領域追求的發展重點之一，在自動化決策和機器學習的幫助下，人工智慧系統可節省大量人力並大幅提高作業效率，金融領域中的信貸服務也可藉著人工智慧系統以提供更便利的使用者體驗。但本文針對人工智慧系統在信貸服務中的應用，發現人工智慧系統用於信用評估的公平性問題，故希望透過檢視反歧視法的規範體系、人工智慧的公平性問題到比較法研究，觀察我國金融法制上的反歧視規範及人工智慧信貸歧視規範，提出立法與監管的建議。

金融業運用人工智慧技術時，為避免公平性疑慮，本應在風險發生前持續保持人工智慧工具的透明性並清楚解釋其決策，同時應確保使用資料及模型穩健且合理，避免不當資料影響了演算法決策。但人工智慧信貸歧視議題實涉及金融機構追求商業利益自由與借款人不受歧視間的價值取捨，判斷的標準隨著法規範追求的目的而有所浮動，信貸作為影響人民生活重要的金融服務，為兼顧金融機構的商業利益下應追求分配正義，人工智慧信貸歧視規範的核心宗旨應是保障受保護特徵群體在信貸審核流程中受有公平對待，不因其具有受保護特徵而受有信用評估或信貸條件上的不利益。然而，欲在人工智慧系統應用的程序中達到公平性的目的，現實上難以透過法律事前規範清楚界線，必須透過結果導向的測試方法，採取事後監管的方法。而為達成結果測試的監理方式，監理政策上需先確立受保護特徵，因此須先釐清基本的歧視定義、受保護特徵的擇定與立法規範模式。

我國目前尚無完整的金融反歧視規範，目前金融規範中多係以公平待客為原則，保障高齡者或身心障礙者。縱使有部分規範得加以解釋沿用，但仍有立法架構零散、欠缺體系且實質規範內容及範圍欠缺明確之問題[306]。相對而言，比較法研究上，從美

[306] 劉祐君（2020），〈我國信貸反歧視規範初探－以美國《公平信貸機會法》為比較〉，233 期，《萬國法律》，頁 96。

doi:10.6342/NTU202500326

國 ECOA 與 Regulation B 均禁止在信貸領域中對具有法定受保護特徵的群體構成差別對待歧視及差別影響歧視，此為我國法先天不足之處。

但現行金融反歧視規範係以人力提供金融服務為假設前提。人力審核的信用貸款流程中，歧視的發生強調的是「因果關係」，但機器學習的依據是資料與結果間的「相關性」，導致人工智慧系統中的歧視結果其原因並非來自於因果關係，難以透過可預見性而得加以防範。因此，無論是 ECOA 或是美國其他的關於貸款的平等規範（如 FHA），在面對人工智慧系統時，仍有其極限。在舉證責任方面，根據 Griggs 案確立的結果，需由原告先負有舉證責任，包括行為人的歧視行為、原告為受保護群體之人、形成差別影響結果，Ricci 案也指出須達到「有堅強基礎的證據」之證明程度才能證明此差別影響構成歧視。因此，信貸程序中受到人工智慧系統藉由中性特徵而有不利影響的問題，從因果關係認定與舉證責任分配上，似難以參酌 ECOA 的差別影響歧視理論解決。

然而，在面對人工智慧信貸歧視時，如本文認為可以透過可解釋人工智慧的運用，識別出演算法偏差，進而透過要求金融機構應負有尋求較小歧視的替代方案的責任，即可以避免受歧視者為證明金融機構以受保護特徵為歧視性影響時的舉證困難，以及避免非受保護特徵的新形態歧視的發生，也可作為監管機關檢查金融機構是否使用人工智慧系統歧視的管道之一。如前揭第二章第二節第一項各款中提出的 Wells Fargo 房貸信用評分歧視案、海軍聯邦信用社歧視案及 Apple Card 案例，都是在應用機器學習的信用評估及決策程序中產生歧視的不公平結果，透過要求金融機構以可解釋人工智慧識別其所使用的人工智慧信用評估系統是否具有歧視後，並要求其負有尋求較小歧視替代方案的責任，得以避免現行法面對人工智慧代理歧視的限制，而解決人工智慧信貸歧視的困境。

本文透過美國公平信貸法及其相關反歧視規範、人工智慧相關規範及學者文獻的研究，探討了金融機構運用人工智慧系統的信貸歧視問題，最後梳理我國現行的金融

doi:10.6342/NTU202500326

反歧視規範與金融業運用人工智慧系統的相關規範，並在 2024 年 3 月中華民國銀行商業同業公會全國聯合會訂定《金融機構運用人工智慧技術作業規範》及同年 6 月金管會發布的《金融業運用人工智慧（AI）指引》基礎下，從我國法制現況出發，對於我國制定一般性金融反歧視規範、具體應指定的受保護特徵、規範效力、規範對象、規範內容以及監管機關權限提出具體建議，並於規範中明定要求金融機構負有選擇較小歧視方案的規範依據，配合可解釋人工智慧的使用，使金融機構得以識別其所使用的人工智慧系統是否帶有歧視，同時持續自我監督。希冀透過完善的立法解決人工智慧系統所帶來的公平性疑慮，在日新月異的未來中，每個人都能夠公平地享受科技帶來的便利服務，自由地追求理想的生活。

doi:10.6342/NTU202500326

參考文獻

壹、中文文獻

（一） 專書

王澤鑑（1999），《法律思維與民法實例》，臺北：自版。

司法院（2011），《美國聯邦最高法院憲法判決選譯》，臺北：司法院。

許志雄（2016），《人權論──現代與近代的交會》，臺北：元照。

（二） 期刊論文

陳凱迪（2018），〈人工智慧發展對金融業之衝擊與因應〉，《財金資訊季刊》，93期，頁 14-22。

陳靜慧（2017），〈歐洲人權法院及歐洲法院對間接歧視之適用〉，《憲法解釋之理論與實務》，9 期，頁 385-436。

焦興鎧（1990），〈美國勞工法上就業歧視之最新發展〉，《美國研究》，20 期，頁 51-116。

焦興鎧（1995），〈勞工法與勞工權利之保障〉，《美國勞工法論文集（一）》，頁 329-364。

黃昭元（2012），〈論差別影響歧視與差別對待歧視之關係－評美國最高法院 Ricci v. DeStefano(2009)判決〉，《中研院法學期刊》，11 期，頁 1-63。

葉啟洲（2014），〈民事交易關係上之反歧視原則－德國一般平等待遇法之借鏡〉，《東吳法律學報》，26 期，頁 143-200。

廖淑君（2012），〈人工智慧與普惠金融－淺析演算法於徵信／授信應用之金融消費者保護議題〉，《財金法學研究》，5 卷，1 期，頁 115-150。

劉祐君（2020），〈我國信貸反歧視規範初探－以美國《公平信貸機會法》為比較〉，《萬國法律》，233 期，頁 91-99。

doi:10.6342/NTU202500326

謝棋楠（2008），〈由美國差別影響歧視制度評台北地院九十一年重勞訴字第五號判決〉，《全國律師月刊》，12 卷 4 期，頁 14-35。

（三） 學位論文

李婷婷（2019），《小額消費信貸信用評分模型的建置–以台灣某銀行為例》，國立清華大學財務金融碩士在職專班碩士論文。

康建民（2008），《個人貸款信用評分模型-以儲蓄互助社為例》，國立暨南大學財務金融學系碩士論文。

楊舒嵐（2010），《憲法價值於私法關係之體現—以隱私權與表現自由之衝突與調和為例》，國立臺灣大學碩士論文。

（四） 決議、函令

金融監督管理委員會 113 年 3 月 14 日金管銀國字第 1120237675 號函

金融監督管理委員會 106 年 6 月 3 日金管銀國字第 10620002360 號函

（五） 政府文件

金融監督管理委員會（2023），〈金融業運用人工智慧(AI)之核心原則與相關推動政策〉。

金融監督管理委員會（2024），〈金融業運用人工智慧(AI)指引〉。

（六） 網路資料

Medium，〈可解釋 AI (XAI) 系列 — SHAP〉，載於：https://medium.com/ai-academy-taiwan/%E5%8F%AF%E8%A7%A3%E9%87%8B-ai-xai-%E7%B3%BB%E5%88%97-shap-2c600b4bdc9e。

Medium，〈平等、公平、現實、正義(解放)之間的差異〉，載於：https://medium.com/vincent-chen/%E5%B9%B3%E7%AD%89-

doi:10.6342/NTU202500326

%E5%85%AC%E5%B9%B3-%E7%8F%BE%E5%AF%A6-%E6%AD%A3%E7%BE%A9-%E8%A7%A3%E6%94%BE-%E4%B9%8B%E9%96%93%E7%9A%84%E5%B7%AE%E7%95%B0-79f7043e8d6e。

中央廣播電臺（2024/06/20），〈台灣金融業使用 AI 占比近 3 成金管會發布指引〉，載於：https://www.rti.org.tw/news/view/id/2210313。

中時新聞網（07/28/2023），〈合庫銀將推「速速貸線上信貸」最快 30 分鐘入帳〉，載於：https://ynews.page.link/Jn8z5。

台灣好報（09/06/2023），〈廈門銀行推出「臺 e 貸」 大陸首款臺胞專屬線上信用貸款產品〉，載於：https://tw.news.yahoo.com/%E5%BB%88%E9%96%80%E9%8A%80%E8%A1%8C%E6%8E%A8%E5%87%BA-%E8%87%BAe%E8%B2%B8-%E5%A4%A7%E9%99%B8%E9%A6%96%E6%AC%BE%E8%87%BA%E8%83%9E%E5%B0%88%E5%B1%AC%E7%B7%9A%E4%B8%8A%E4%BF%A1%E7%94%A8%E8%B2%B8%E6%AC%BE%E7%94%A2%E5%93%81-083204181.html。

金管會（2023/11/28），〈金管會就金融業運用人工智慧(AI)指引草案公開徵詢外界意見〉，載於：https://www.fsc.gov.tw/ch/home.jsp?id=96&parentpath=0,2&mcustomize=news_view.jsp&dataserno=202312280004&dtable=News。

科技大觀園，〈電機資工的現況與未來：生活中的演算法〉，載於：https://scitechvista.nat.gov.tw/Article/c000003/detail?ID=084a957b-aa02-4833-a08d-ec0788e97d50。

財訊（09/29/2021），〈傳統銀行程序冗長 自動化借貸系統有創舉 AI 助攻信貸平台股價暴衝 5 倍〉，載於：https://www.wealth.com.tw/articles/9e9060cd-9248-499e-801e-f1744f944c5d。

doi:10.6342/NTU202500326

財團法人資訊工業策進會數位轉型研究院，〈金管會AI指引、銀行公會AI規範發布，三大重點一次看懂！〉，載於：https://www.find.org.tw/index/indus_trend/browse/875b2d221ea9f38c4fc418f3d0089ead/。

凱基銀行（09/27/2024），〈信用貸款是什麼？信貸流程有哪些？貸款申請要多久？貸款流程申請必備資訊總整理！〉，載於： https://www.kgibank.com.tw/zh-tw/kgibarticleshome/loan/pl-process。

貳、英文文獻

（一） 專書

Becker, G. S. (1971), The Economics of Discrimination, University Of Chicago Press.

Bentham, Jeremy (revised 2017), An Introduction to the Principles of Morals and Legislation, https://www.earlymoderntexts.com/assets/pdfs/bentham1780.pdf.

Fredman, Sandra (2011), Discrimination Law, Oxford University Press.

Hellman, Deborah & Moreau, Sophia, eds. (2013), Philosophical Foundations of Discrimination Law, Suffolk University Law School Research Paper.

Lauer, Josh (2017), Creditworthy, Columbia University Press.

Lippert-Rasmussen, Kasper (2014), Born Free And Equal?: A Philosophical Inquiry Into The Nature Of Discrimination, Oxford University Press.

Rawls, John (1971), A Theory of Justice, Harvard University Press.

Sandel, Michael J. (2010), Justice: What's the Right Thing to Do?, Farrar Straus & Giroux.

（二） 期刊論文

Andreeva, Galina, Ansell, Jake & Crook, Jonathan (2004), Impact of Anti-Discrimination Laws on Credit Scoring, 9 J. Financ. Serv. Mark 22.

doi:10.6342/NTU202500326

Black, Emily et al. (2023), Less Discriminatory Algorithms, 101 Georgetown Law Journal 53, https://www.law.georgetown.edu/georgetown-law-journal/wp-content/uploads/sites/26/2024/11/Black-et-al._LessDiscriminatory.pdf.

Collins, Hugh (2003), Discrimination, equality and social inclusion, The Modern Law Review 66.

Deeks, Ashley (2019), The Judicial Demand for Explainable Artificial Intelligence, 119 Columbia Law Rev.1829.

Di Palma, M., Carioti, D., Arcangeli, E., et al. (2022), The biased hand. Mouse-tracking metrics to examine the conflict processing in a race-implicit association test, 17(7) Plos One 2, https://pmc.ncbi.nlm.nih.gov/articles/PMC9328548/.

Fuster, Andreas, Goldsmith-Pinkham, Paul S., Ramadorai, Tarun & Walther, Ansgar (2022), Predictably Unequal? The Effects of Machine Learning on Credit Markets, 77 Journal Of Finance 5, http://dx.doi.org/10.2139/ssrn.3072038.

Garcia, Ana Cristina Bicharra, Garcia, Marcio Gomes Pinto & Rigobon, Roberto (2023), Algorithmic Discrimination in the Credit Domain: What Do We Know about It?, Ai & Society, https://doi.org/10.1007/s00146-023-01676-3.

Gerards, Janneke & Zuiderveen Borgesius, Frederik (2020), Protected Grounds and the System of Non-Discrimination Law in the Context of Algorithmic Decision-Making and Artificial Intelligence, 20 Colo. Tech. L.J. 1, https://ctlj.colorado.edu/wp-content/uploads/2022/06/PROTECTED-GROUNDS-AND-THE-SYSTEM-OF-NON-DISCRIMINATION-LAW-IN-THE-CONTEXT-OF-ALGORITHMIC-DECISION-MAKING-AND-ARTIFICIAL-INTELLIGENCE.pdf.

Heringa, Aalt Willem (1999), Standards of Review for Discrimination: The Scope of Review by the Courts, Non-Discrimination Law: Comparative Perspectives 25.

doi:10.6342/NTU202500326

King, Allan G. & Wang, Alice H., Artificial Intelligence As A Less Discriminatory Alternative, 33 Public Policy 375 (2023), https://scholarship.law.ufl.edu/jlpp/vol33/iss3/2/.

Lawrence, Charles R. III (1987), The Id, the Ego, and Equal Protection: Reckoning with Unconscious Racism, 39 Stan. L. Rev. 317.

Lepri, Bruno et al. (2018), Fair, Transparent, and Accountable Algorithmic Decision-Making Processes: The Premise, the Proposed Solutions, and the Open Challenges, 31 Philos. Technol. 611.

McGinley, Ann C. (2010), Discrimination Redefined, 75 Mo. L. Rev 443.

Odinet, Christopher K. (2022), Fintech Credit and the Financial Risk of AI, U. Iowa Legal Stud. Research Paper, No. 2021-39.

Selbst, Andrew D. & Barocas, Solon (2018), The Intuitive Appeal of Explainable Machines, 87 Fordham L. Rev. 1085, https://ir.lawnet.fordham.edu/cgi/viewcontent.cgi?article=5569&context=flr.

Wasserman, David, Discrimination, concept of (1998), 1 Encyclopedia Of Applied Ethics, 807.

Wu, Jason Jia-Xi (2024), Algorithmic Fairness in Consumer Credit Underwriting: Towards a "Harm-Based" Framework for AI Fair Lending, 21 Berkeley Business Law Journal 1, https://doi.org/10.15779/Z38CF9J785.

Zuiderveen Borgesius, Frederik J. (2020), Strengthening Legal Protection against Discrimination by Algorithms and Artificial Intelligence, 24 Int. J. Hum. Rights 1572.

（三） 判決案例

A.B. & S. Auto Service, Inc., v. South Shore Bank of Chicago, 962 F.Supp. 1056, 1061 (N.D. Ill. 1997)

Albemarle Paper Co. v. Moody, 422 U.S. 405 (1975)

Braxton v. Wells Fargo Bank, N.A. a Delaware corporation, 3:22-cv-01748, (N.D. Cal.)

doi:10.6342/NTU202500326

DeJung v. Superior Court, 169 Cal.App.4th 533, 87 Cal. Rptr. 3d 99 (Cal. Ct. App. 2008)

Federal Trade Commission, V. Passport Automotive Group, Inc. No. 8:2022cv02670 (US District Court for the District of Maryland)

Griggs v. Duke Power Co., 401 U.S. 424, 91 S. Ct. 849 (1971)

Lewis v. ACB Bus. Services, Inc., 135 F.3d 389, 406 (6th Cir. 1998)

Monson v. Rochester Athletic Club, 759 N.W.2d 60 (Minn. Ct. App. 2009)

Ricci v. Destefano, 530 F.3d 88 (2d Cir. 2008)

Union of Refugee Women v. Director, Private Security Industry Regulatory Authority (CCT 39/06) [2006] Z.A.C.C. 23, paras 122–123 (S. Afr.).

（四） 研究報告

Consumer Financial Protection Bureau (2024), Fair Lending Report of the Consumer Financial Protection Bureau, , https://files.consumerfinance.gov/f/documents/cfpb_fair-lending-report_fy-2023.pdf.

European Banking Federation (2019), EBF Position Paper on AI in the Banking Industry, https://www.ebf.eu/wp-content/uploads/2020/03/EBF-AI-paper-_final-.pdf.

Federal Trade Commission, Discrimination And Bias In Automated Systems, https://www.ftc.gov/system/files/ftc_gov/pdf/EEOC-CRT-FTC-CFPB-AI-Joint-Statement%28final%29.pdf.

Financial Stability Board (2022), Fintech and Market Structure in the COVID-19 Pandemic, https://www.fsb.org/wp-content/uploads/P210322.pdf.

OECD (2019), Scoping the OECD AI principles, OECD Digital Economy Papers, https://www.oecd-ilibrary.org/docserver/d62f618a-en.pdf?expires=1589038170&id=id&accname=guest&checksum=4E099F3617BD00290FCC34E651F0920A.

doi:10.6342/NTU202500326

OECD (2020), OECD Business and Finance Outlook 2020: Sustainable and Resilient Finance, OECD Publishing, Paris, https://doi.org/10.1787/eb61fd29-en.

OECD (2022), OECD Framework for the Classification of AI systems, OECD Digital Economy Papers, https://doi.org/10.1787/cb6d9eca-en.

U.S. Equal Employment Opportunity Commission, Select Issues: Assessing Adverse Impact in Software, Algorithms, and Artificial Intelligence Used in Employment Selection Procedures Under Title VII of the Civil Rights Act of 1964, https://www.eeoc.gov/laws/guidance/select-issues-assessing-adverse-impact-software-algorithms-and-artificial#_ednref21:~:text=What%20is%20the%20%E2%80%9Cfour%2Dfifths%20rule%E2%80%9D%3F.

U.S. Public Interest Research Group (USPIRG) and the Center for Digital Democracy (CDD) (2015), Expanding Access to Credit through Online Marketplace Lending, Comments To The U.S. Department Of The Treasury, https://www.democraticmedia.org/sites/default/files/field/public/2015/uspirg_cdd_marketplacelendingrf_fnal30sept2015.pdf.

（五） 網路資料

Aba Banking Journal, Fed’s Barr highlights digital redlining risk resulting from AI tools, https://bankingjournal.aba.com/2023/07/feds-barr-highlights-digital-redlining-risk-resulting-from-ai-tools/.

AI Incident Database, Incident 92: Apple Card's Credit Assessment Algorithm Allegedly Discriminated against Women (2019), https://incidentdatabase.ai/cite/92/.

doi:10.6342/NTU202500326

Bet, Black Home Mortgage Borrowers Sue Navy Federal Credit Union For Discrimination (2023), https://www.bet.com/article/vlo62w/navy-federal-credit-union-lawsuit-racial-discrimination.

Bloomberg, Artificial Intelligence Has Some Explaining to Do (Dec 12, 2018), https://www.bloomberg.com/news/articles/2018-12-12/artificial-intelligence-has-some-explaining-to-do.

Bloomberg, Wells Fargo Rejected Half Its Black Applicants in Mortgage Refinancing Boom (Mar. 11, 2022), https://www.bloomberg.com/graphics/2022-wells-fargo-black-home-loan-refinancing/?utm_source=website&utm_medium=share&utm_campaign=copy.

CNN, The Nation's Largest Credit Union Rejected More Than Half Its Black Conventional Mortgage Applicants (Dec. 14, 2023), https://www.cnn.com/2023/12/14/business/navy-federal-credit-union-black-applicants-invs.

Consumer Federation of America, Consumer Groups Call on CFPB to Protect Consumers from Discriminatory Algorithms Used by Banks and Other Financial Institutions to Make Credit Decision (June 26, 2024), https://consumerfed.org/press_release/consumer-groups-call-on-cfpb-to-protect-consumers-from-discriminatory-algorithms-used-by-banks-and-other-financial-institutions-to-make-credit-decisions/.

Cornell University, The Fairness of Credit Scoring Models (2024), http://arxiv.org/abs/2205.10200.

Datenschutz-notizen, Privacy and AI: Schufa algorithm condemned by the CJEU, https://www.datenschutz-notizen.de/privacy-and-ai-schufa-algorithm-condemned-by-the-cjeu-0946847/.

Federal Trade Commission, Federal Trade Commission Takes Action Against Passport Automotive Group for Illegally Charging Junk Fees and Discriminating Against Black and Latino Customers (October 18, 2022), https://www.ftc.gov/news-events/news/press-

doi:10.6342/NTU202500326

releases/2022/10/federal-trade-commission-takes-action-against-passport-automotive-group-illegally-charging-junk-fees.

Forbes, What Is An Algorithm? Defining And Applying Algorithms, https://www.forbes.com/sites/technology/article/what-is-an-algorithm/

Geeksforgeeks, Supervised and Unsupervised Learning (23 Sep, 2024), https://geeksforgeeks.org/supervised-unsupervised-learning/.

Georgetown Online Public Health, Equity vs. Equality: What's the Difference? (November 5, 2020), https://onlinepublichealth.gwu.edu/resources/equity-vs-equality/.

Github, Counterfactual Explanations, https://christophm.github.io/interpretable-ml-book/counterfactual.html.

Github, Local Surrogate (LIME), https://christophm.github.io/interpretable-ml-book/lime.html.

Github, Permutation Feature Importance, https://christophm.github.io/interpretable-ml-book/feature-importance.html.

Housing Wire, Mortgage Underwriting Algorithm at Heart of Wells Fargo's Racial Disparity Lawsuit (June 14, 2024), https://finance.yahoo.com/news/mortgage-underwriting-algorithm-heart-wells-161043355.html.

Investopedia, Effects Test: What It Means, How It Works (2022), https://www.investopedia.com/terms/e/effects-test.asp.

Investopedia, What Is Redlining? Definition, Legality, and Effects, https://www.investopedia.com/terms/r/redlining.asp.

JMP, Introduction to machine learning: easy-to-understand interpretation of decision trees (May 20, 2024), https://community.jmp.com/t5/JMP-Blog/%E6%A9%9F%E5%99%A8%E5%AD%B8%E7%BF%92%E5%85%A5%E9%96%80-

doi:10.6342/NTU202500326

%E9%80%9A%E4%BF%97%E6%98%93%E6%87%82%E8%A7%A3%E8%AE%80%E6%B1%BA%E7%AD%96%E6%A8%B9/ba-p/753561.

Knowledge, How Anti-Discriminatory Measures Can Worsen AI Bias (2023), https://knowledge.insead.edu/operations/how-anti-discriminatory-measures-can-worsen-ai-bias.

Machine Learning, Linear regression, https://developers.google.com/machine-learning/crash-course/linear-regression.

Medium, A Quick Guide to Boosting in ML (Mar. 21, 2018),https://medium.com/greyatom/a-quick-guide-to-boosting-in-ml-acf7c1585cb5/.

Medium, Understanding the Essentials of Model Distillation in AI (Jun 8, 2024), https://medium.com/stream-zero/understanding-the-essentials-of-model-distillation-in-ai-1e97403bee8a.

Merriam-Wester.Com, Algorithm, https://www.merriam-webster.com/dictionary/algorithm.

MIT Management, Machine learning, explained (Apr 21, 2021), https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained.

Mit Tech. Rev., Is AI Riding a One-Trick Pony? (2017), https://www.technologyreview.com/s/608911/is-ai-riding-a-one-trick-pony/.

New York Institute of Finance, The C's and P's of Credit Risk Analysis, https://info.nyif.com/do-know-your-cs-and-ps-for-credit-analysis/.

New York State Department of Financial Services, Report on Apple Card Investigation (2021), https://www.dfs.ny.gov/system/files/documents/2021/03/rpt_202103_apple_card_investigation.pdf.

Office of the comptroller of the Currency, Comptroller's Handbook: Fair Lending (2023), https://www.occ.treas.gov/publications-and-resources/publications/comptrollers-handbook/files/fair-lending/index-fair-lending.html.

doi:10.6342/NTU202500326

Oracle Ai &Data Science Blog, Using the Artificial Neural Network for Credit Risk Management (Jan. 23, 2019), https://blogs.oracle.com/ai-and-datascience/post/using-the-artificial-neural-network-for-credit-risk-management.

Reuters, Insight - Amazon scraps secret AI recruiting tool that showed bias against women (October 11, 2018), https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/.

SpiceWorks, What Is Logistic Regression? Equation, Assumptions, Types, and Best Practices (April 8, 2022), https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-logistic-regression/.

The Regulatory Review, How the FTC Could Regulate Algorithmic Discrimination (Jan 25, 2024), https://www.theregreview.org/2024/01/25/shaikh-how-the-ftc-could-regulate-algorithmic-discrimination/.

U.S. Department of Justice, Federal Coordination and Compliance Title VI Manual Section VII- Proving Discrimination- Disparate Impact, https://www.justice.gov/crt/fcs/T6Manual7#P.

United States Consumer Financial Protection Bureau & Department of Justice, Joint Statement On Enforcement Efforts Against, https://www.ftc.gov/system/files/ftc_gov/pdf/EEOC-CRT-FTC-CFPB-AI-Joint-Statement%28final%29.pdf.

World Economic Forum, International Women's Day: What's the difference between equity and equality? (Mar 3, 2023), https://www.weforum.org/stories/2023/03/equity-equality-women-iwd/.

doi:10.6342/NTU202500326