國立臺灣大學法律學院法律學系

碩士論文

Department of Law
College of Law
National Taiwan University
Master Thesis

公部門中的人工智慧
—人為介入作為正當使用人工智慧的必要條件

Artificial Intelligence in the Public Sector: Human Intervention as the Necessary Condition for Legitimate Use of Artificial Intelligence

呂胤慶

Yinn-Ching Lu

指導教授：林子儀博士、蘇慧婕博士

Advisors: Tzu-Yi Lin, J.S.D., & Hui-Chieh Su, Dr.iur

中華民國 110 年 10 月
October 2021

doi:10.6342/NTU202103631

doi:10.6342/NTU202103631

國立臺灣大學碩士學位論文
口試委員會審定書

公部門中的人工智慧
—人為介入作為正當使用人工智慧的必要條件
Artificial Intelligence in the Public Sector:
Human Intervention as the Necessary Condition for
Legitimate Use of Artificial Intelligence

本論文係呂胤慶君（學號：R07A21014）在國立臺灣大學法律學系完成之碩士學位論文，於民國110年10月4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：林子儀、薛慧敏

口試委員：劉靜怡、薛慧敏

林子儀、邱文聰

doi:10.6342/NTU202103631

誌謝

我想先謝謝妳[1]願意在圖書館的書架前駐足，抽出這本不起眼的論文。抑或是，妳是在網路搜尋頁面中，自願的選擇下載了這本標題有點長的論文。我不太確定妳是抱著什麼樣的心態打開這本論文的。不過，我想告訴妳，我即將要從這個待了三年多的地方畢業，去下一個地方。我想在這邊，用這本論文應該會被最多人次閱讀的角落，跟妳分享一個故事。這個故事關於這本論文背後的世界，關於一些這三年中遇到一些美好，一群很認真、很熱情的人們。這個故事可能會有點長，角色可能有些多，所以我希望妳在閱讀這個部分之前，可以先喝幾口左手邊的咖啡。衛生紙倒是不必準備，但如果妳跟我一樣過敏的話，還是放在右手邊比較好。

這本論文的構想起源於西元（下同）2019 年 10 月。那個時候，我真的對於使用很久的社群媒體又愛又恨。喜歡他帶給我的人際關係，但又討厭他帶給我的內褲廣告。那個時候大家都在討論怎麼去管制社群媒體，在好幾位厲害的學長姐都把這個題目挖遍後，我只好作罷。不過在發現社群媒體產生問題的原因，是資料以及演算法造成時，而且這種應用方式也很常用在行政管制，我就決定深入演算法以及資料，釐清它們造成的其他影響。

這本論文寫了整整兩年。我這本論文初稿的第一版完成於 2021 年 6 月 24 日。可能是因為悶在家裡寫了太久吧，我不僅寫的文風不順，還牛頭不對馬嘴[2]，好多人都說看不懂我在寫什麼。所以，我又整本刪掉，重新寫了一遍。妳知道我那個時候在想什麼嗎？我其實有點忘記了。不過我在改寫的過程中，其實覺得蠻踏實的，因為好多人都在幫我，讓我沒有後顧之憂。最後，終於在 10 月 4 日的時候，從霖澤館實習法庭後的第四會議室完成了將近四小時的口試。妳知道四小時是什麼意思嗎？法學院碩士班公法組的課，一個人有一小時的時間口頭報告加討論就已經很多了。還好我平常熱能攝取充足，當天

[1] 本部分第二人稱皆以「妳」作為稱呼。

[2] 牛頭不對馬嘴是我說的。蘇老師是用「你把一本論文寫成兩本論文」非常不傷人的方式道出了這個事實。

doi:10.6342/NTU202103631

結束之後也不過是喉嚨有點沙啞而已。花三週的時間，我又改了這本論文大概一半的論述，最後在 10 月 26 日定稿。[3]

***

妳以為這個故事中只有我這個角色嗎？大錯特錯。我只是負責把這個故事寫出來的人而已。所以，下面我想跟你介紹這個故事中的角色。我之所以要跟妳介紹他（她）們是因為：**我超級感謝他（她）們的**。

首先，是我的兩位指導教授，林子儀老師與蘇慧婕老師。

我絕對不會跟妳說什麼「林老師很帥」這種有眼睛就會發現的事實。我想跟妳分享的是林老師對我（論文）的堅持與期許。林老師對於行文的「用語」有特殊的堅持。我想是因為他知道文章寫出來就是會被別人讀的。他不希望論文的內容因為不恰當的用語而造成他人的誤解。[4]林老師影響我的人生哲學是：事情，要嘛不做，要做就到最好。論文的撰寫過程中，我總是持續的往這個目標邁進。林老師指導我論文的過程中，總是以「收穫最多」作為指導的首要指標：在選擇題目的時候希望選擇對自己未來最有幫助的論文題目；在論證上希望我採取對於社會最有實際意義的討論方向；在時間限度內為我保留最多的準備時間。林老師對於我論文，直到最後一個版次、最後一天，都沒有鬆懈，逐字斧正，嚴格把關。[5]此外，林老師總是希望我能夠兼顧理論與實際，要提出現實中具有可行性、並且是有意義的解方，而不要淪為

[3] 這邊可以為妳補充一下這條時間軸背後的進行式。地球在2019年年底的時候，受到一種強韌病毒的威脅。最初發現於中國武漢，而後傳播至全球，災難性的改變了這個社會。我的國家，臺灣，作為這個世界的一份子，也沒有辦法倖免於這場瘟疫。從最初遊刃有餘的抵擋，到後來的措手不及。首都臺北市與新北市於2021年5月15日進入嚴苛的第三級防疫警戒，四天後這個警戒遍及全國。為了生存，人們的生活模式發生改變：人手一個口罩，遠離實體中人與人的接觸、仰賴虛擬上人與人的連結。對於我這個法學院學生而言，上一週還在法學院的走廊上打鬧，下一週就開始測試哪一個視訊會議軟體不會斷線。對了，妳打疫苗了嗎？妳那裡還需要打疫苗嗎？

[4] 這邊可以說一下這本論文中林老師建議的幾個用語。第一個應該是大家很熟悉的是「technology」的用語。林老師認為應該翻譯為「技術」，而非科技，理由可以參見我第一章的註腳27。第二個是「automated decision-making」。這個中文用語主要是因為文義脈絡的影響。我一開始是翻成「自動化決策」。但「決策」這個用語本身有一些實踐性的政策考量，但我的文中其實要講的其實只是適用法律後的結果，所以老師建議我翻譯為「自動化決定」。第三個是「explanation」，我一開始是翻譯為「解釋」，而林老師建議我翻譯為「說明」。這個是因為主要是考量到特定規範內容的理解，理由請參見第一章的註腳19。

[5] 林老師甚至在疫苗接種第二劑副作用的體驗過程中，一字一句的讀完我的第一章，並且提出意見。我這個學生的真的是只能感謝老師。

doi:10.6342/NTU202103631

蛋頭研究生。即便每次跟林老師的討論都充滿了爭執以及異見，但林老師永遠都保留給我最大的學術自由，希望能透過刺激，促使我產出更有價值的知識。

其次，我要跟妳介紹永遠都知道我在想什麼的蘇老師，以及教我論證的蘇老師。從基本人權專題、法學德文名著選讀、大一憲法課，蘇老師都是站在培養學生自己解決問題、提出意見的能力的觀點來規劃課程。從文本的分析、論證的重新拆解、找出論點，到有道理的支持/反駁特定論點；甚至是從文本中如何抽取知識，建構自己的思考地圖。作為老師的助理，老師無私的讓我透過觀察學術後臺，教我如何從事有效率的學術研究。作為老師的學生，除了論文的指導之外，老師提供各種即時的協助，不論是知識取得的困境、心理煎熬的調適。此外，老師透過「誘導式」的指導方式，告訴我問題解決的方式，但不直接告訴我答案，讓我能夠獨自享受研究中的樂趣以及成就感。我想如果我在這條路上感覺到任何的樂趣，我都應該歸功於蘇老師，因為是蘇老師教我如何享受研究的過程。

另外，這本論文能夠變成今天這個樣子[6]，我要跟妳介紹另外兩個老師，分別是劉靜怡老師以及邱文聰老師。這兩位老師很仔細的閱讀我不好的中文、近乎通靈式的理解我想表達的主張。在口試中，二位老師梳理我的推論過程、一一檢視支撐論點的理據。最後，依序請我就內容進行澄清，並提出論點的挑戰。二位老師對我論文的重新解構，讓我清楚的知道論證上的瑕疵以及不足之處。[7]雖然改寫時間有限，但我盡力讓口試中的知識交流，呈現在最終的版本之中。

除了擔任口試委員之外，劉靜怡老師也讓我長期擔任助理的工作。不僅訓練我的學術能力以及讓我有機會參與許多有意義的活動，像是去中研院法律所資訊法中心觀賞各位老師神仙打架，或者是協助促成憲政時刻的來臨。

[6] 我指的並不是內容多有學術貢獻，而是口試本跟論文定稿的版本有十分大的差異。當然，我只能將此歸功於兩位口試委員在口試中的各種意見與建議。講一個比較不相干的。有一次蘇老師在收到我的論文初稿N版後3天，對我說出了一個評價：「你是健達出奇蛋阿，每次的東西都長不一樣」。我始終不知道這個類比推論是褒還是貶。不過我蠻喜歡這個類比的。

[7] 邱老師要開始問我問題時，甚至拿出隨身碟說「胤慶，你放一下裡面標題是GDPR22的投影片」。我真的是感激涕零。老師為了讓我清楚的知道論證在哪裡出問題，甚至還做了投影片。

doi:10.6342/NTU202103631

當劉老師助理的過程中，不僅在文獻的閱讀能力上有所提升，每次跟劉老師聊天完，我都覺得我的「視野」都有顯著的長進，知道為什麼今天會有這個投書、昨天為什麼有這篇報導。劉老師精準的眼光，都讓我對於「社會脈絡」有更清楚的觀察。

到資訊法中心參加讀書會後，才真的認識了邱文聰老師。邱老師真的是反應快、說話清楚、充滿洞見，而且沒有架子。對於科學、技術的概念，或抽象的法律論證推導，邱老師都能用最清楚、最白話的言語加以說明。每次聽邱老師的報告或者是知識分享，我都又愛又懼。愛的是邱老師又讓我的視野更寬廣、想法更清晰；懼的是邱老師即將限制我的發展方向，因為邱老師的想法以及創見總是根基紮實，讓我沒有辦法推翻。如妳所見，我的論文就是沿著邱老師的創見，繼續發展了（攤手）。我還要感謝邱老師分享他未出版的手稿讓我在論文中加以引用。

除了指導教授與口試委員之外，我還要跟妳介紹其它啟發我很多的老師。

李建良老師每次都全方位的回答我所有的問題，並且關心我的人生近況。我一開始拎著國考考題去找老師問爭點、之後在課堂上帶著判決去找老師問問題意識，到後續論文撰寫中跟李老師討論歐盟資料保護法的研究方向。李老師的回答總是超乎我小腦袋的想像。李老師會先告訴我特定問題背後的關懷，然後跟我說明問題在思考上的推演，最後告訴我回答可能的走向。李老師每次的回答，都讓我對於不同的問題有更全面的理解。此外，我也感謝李老師每次見到我都會關心我的研究，以及人生進度。

林明昕老師經常關心我，不論課堂報告、法學德文、生活日常或者論文進度。除了驚訝於林老師對於人際、社交網絡的全面掌握之外，每次跟林老師參與會議、討論大小問題，都好佩服林老師的敏捷思考，以及機智反應。

孫迺翊老師在初來乍到臺大的時候，就讓我這個初心（粗心）研究生長期的擔任老師的助理。孫老師做事總是親力親為，不論是上課、從事研究，或是課外活動，老師都是燃燒自己的全力付出。每次跟老師討論法律問題、參與憲法時刻，都會發現老師背後的熱情。還有，是孫老師在我心中種下歐盟法種子的。

doi:10.6342/NTU202103631

黃昭元老師在憲法專題研究課程中對於我課堂的提問，產生了許多的外溢作用，例如在其它場合的問題回應，或者我論文實際論證的發想。此外，黃老師給我的機會、對我無限包容，是我在論文撰寫中最大的感謝之一。

另外，不論是在課堂中，或是擔任助理的過程中，我還遇到了好多人很好的老師，像是王立達老師、吳全峰老師、胡博硯老師、范文清老師、范秀羽老師、翁岳生老師、張文貞老師、莊世同老師、許宗力老師、陳柏良老師、陳舜伶老師、程明修老師、黃丞儀老師、黃詩淳老師與蘇彥圖老師。這些老師真的都幫了我好多，不論是助理工作機會的給予、學術能力的培養、文章撰寫的建議，或人生近況的關心。

我在寫完論文第二章關於歐盟資料保護法的內容後，在中央研究院歐美研究所 2020 全國研究生歐美研究論文發表會發表。我還記得當時的評論人簡資修老師，以及與會的何之行老師、洪德欽老師、陳弘儒老師與顏厥安老師提供給我好多好多的意見。當天跟這些老師有激烈的討論，直到今天都讓我記憶猶新。

一路走來，我都深深的覺得能夠認識這些老師實在是太好、太幸運了。

***

除了以上的老師之外，我也想跟妳介紹一些我在研究所就讀過程中認識的好朋友們，他/她們對於我的論文都有直接的貢獻。

佳樺在不論在上課的過程中，或是我論文發表的過程中，都認真的對待我的報告，向我提出了許多重要的問題，重演早期的公法組風華。此外，在我跑到中研院玩耍的時候，佳樺也向我分享了許多論文撰寫的心得。

第一次參加的論文發表是國祐的，那也是我嚮往的論文發表[8]。我在國祐身上學習到的是「細膩」，不論是論證上的細膩或者是做事上的細膩。研究所就讀的過程中，我有機會能跟國祐一起共寫報告（李老師的基本權利專題研究）。誠實的說，這次共寫報告是我上研究所生涯中最快樂的合作寫作的經驗。不論是議題的發想、論點的採取或者是報告的呈現，我們都可以理性

[8] 後來也某種程度雷同，我指的是被劭楷問倒的部分。

doi:10.6342/NTU202103631

的互相說服。此外，國祐多次的閱讀我的論文稿件，而且都提出了重要的問題，對於我的論文形成有不可抹滅的貢獻。

雙重同門的劭楷在我碩士班中真的巨大的幫助了我，也影響了我。在學術上，劭楷對於理論的愛好以及精通，直接啟蒙了我，讓我的閱讀視野上，慢慢的移出法釋義學。除此之外，劭楷總是不分場合的給我好多建設性的意見。像是課堂上的意見交換（蘇老師的基本人權專題研究）、報告的犀利評論（黃老師的憲法專題研究）、論文發表會上的問題十連擊（至今仍倒地不起），或在市政府的廁所裡聆聽我第 N 次的報告。劭楷應該是同輩中看我論文最多次、最理解我論點的人。這本論文中，劭楷的貢獻無疑是超級巨大的。不僅如此，劭楷還分享許多資訊與機會給我。我對於劭楷的無私以及熱心，始終銘記在心。

芳晨是我在碩士班認識的第一個公法組同學，除了一起修了好多不同的課外，也一起當同事。芳晨的細心、學術視野、有條理的論述以及柔軟的言語，是我每次都努力很想學，但學不起來的。謝謝芳晨每次都認真對待我的報告、學術展演，並提出好多很有建設性的意見。我每次放在文章中的概念連結、投影片的細微設計都能被芳晨解讀出來。

很強的培琪也是我的同事。她是強者，什麼事情都難不倒她，而且她效率好高，講話不會拐彎抹角，跟她共事很開心。在論文撰寫的後期，培琪幫我看出了一個很關鍵的問題，而這個問題甚至擊倒了我當時論文中的一個重要論證。感謝培琪的細心讓我沒有在口試中出包。此外，培琪二話不說的幫我潤飾英文摘要，我超級感動。

最會比賽的睿恩幫我看了至少三個版本的初稿、聽我無數次的發表，每次都問我很重要的問題，總是在我的重要日程前幫我打氣、給我祝福。此外，睿恩每次用黑魔法（法實證研究）從事的法律研究，並且有條理的與理論、法釋義學加以結合，都給我好多的刺激以及持續前進的動力。

作為一本科學與技術的論文，俊中以及東文兩個剛好對於人工智慧內涵有深入理解的好朋友，提供給我好多的資訊。其中東文跨海協助，提供給我資料產業的第一手資訊，並且幫我評估我的法律解方在實際上的可行性。

doi:10.6342/NTU202103631

隨著朋友們畢業的畢業、工作的工作，逐漸沒有熟識的人。還好以前一起打球的永賦，不僅提供公共論壇供法律子民們交換意見，還沒有二話的就願意協助擔任我論文口試的紀錄，幫我處理口試當天的各種行政庶務，讓我能夠專心的口試。

我是一個要求格式的形式主義者，因此在感性上是沒有辦法容忍自己在文字上的錯誤。但在撰寫論文的過程中，我實在沒有能力看出我論文中的錯字。因此，我想跟妳介紹偉蓁。偉蓁幫我從頭到尾看過論文，幫我挑出這些我不想承認識是我自己寫的錯字與怪句。

妳知道的，我是一個很容易緊張的人，經常因為心情而誤事。還好我有幾個義不容辭的好朋友。詔安、彭敏、杰廷、銘恩、瑞翔、芳晨還有劭楷，他/她們每個人在自己焦頭爛額之際，都撥了至少一個小時的時間給我，為我準備口試，聽我報告以及給我受用的評論。在口試的過程中，我居然能夠在心情不受影響的情形下，跟口試委員來回應答。我想他（她）們的幫助居功厥偉。另外，一直以來的好友之瀚，即便他總是要務在身，但他總把我的事當作自己的事在做，幫我在最後的時限內潤飾了英文的摘要。

除了論文的討論伙伴外，在辛亥監獄的翔甯是我廣義公法問題的討論對象。感謝翔甯持續的與我討論公法寶庫內的寶藏，以及呈現給我司法圈內的有趣現象。

然後，我想跟妳介紹另外幾個跟我同一間研究室的好朋友們，哲偉、瑞翔、王傑與堂俊。上面這四個人的順序，我是依照研究能力高低來排的。他們是我進碩士班前就認識的朋友，像我跟瑞翔是同校 9 年的同學、跟堂俊還有哲偉是同校 6 年的同學，而王傑則是同校 5 年。我很高興有機會能夠跟他們在霖澤館 1811 室討論課業，有時候還會一起研究一些運動法跟電力學的東西。我自知我對於這個領域的研究能力比較不足，幸好他們幾個都很包容我。就讀研究所的過程中，如果沒有他們，我雖然可能會比較快畢業，但我的碩士班生活一定會非常乏味。謝謝他們增添我在臺大讀書過程中的色彩。

最後，我想要跟妳介紹，我這輩子中最重要的幾個人，父親、母親與妹妹。父親與母親用力的把我生下，並扶養我長大。我不是個好養的孩子，從

doi:10.6342/NTU202103631

小就體弱多病，時常讓二人傷透腦筋。我感謝爸媽能夠讓我能夠在碩士生涯中，無憂無慮的做研究，並且無條件的支持我選擇人跡罕至的道路。除此之外，爸爸、媽媽以及妹妹為這本論文的各個版本多番照料。爸爸帶著老花眼鏡幫我抓最後一個版本的錯漏字，看著平常與他所撰寫書狀差異甚遠的文字；媽媽在日常之際，將論文的重點一一讀過；妹妹在多個課堂報告夾攻的情況下，幫我一字一句的讀過，以她的語言學專長挑出所有瑕疵的句子。就此，我可以跟妳說，這本論文的字裡行間都是他（她）們的影子。我這本論文是獻給我家人的。

***

一般而言，誌謝寫到這邊就差不多完結了。不過從以上角色介紹直接進入到我死板板的論文，讀起來可能會有點突兀。我想要在這裡插入一個類似學思歷程的東西，說一說我在這兩年內的思考路徑，以及撞了什麼牆。如果妳對於這類的研究有點興趣的話，或許可以少繞點路。

選定這個「技術與法律」類型的題目後，我主要是從兩個方向著手。一方面，要瞭解「技術（人工智慧）」的內涵，另一方面則是要選擇特定的基準（法律）來針對技術進行評價。對於法律人而言，其實要理解技術內涵要花蠻大的功夫的。此時，若有機會與具有相關技術背景的人作討論，其實對於技術內涵的掌握可更高、更正確。不過，因為技術的進步必然會伴隨著評價內涵的更動。撰寫這種題目，真的是蠻累人的，因為要持續的觀察技術的變遷，而變更評價的結果。

在評價基準的建構上，是我這本論文努力最多的部分。人工智慧作為新的社會事實，目前實際的規範仍舊不太多。在討論上，許多人都是從一些倫理性的要求開始討論。因為，我自己沒什麼能力掌握這種比較抽象的內容，所以我理所當然的從現有的少數實際規範（歐盟資料保護法）切入討論。不過這個規範在現實中仍舊沒有太多的案子，所以仍舊沒有辦法很實質的討論。

論文寫作到這裡的時候，我盤旋了好久，一直下不來。後來，我發現了另外一個討論脈絡，也就是人工智慧對於「法學」本身的影響（國內也以此議題辦了不少場研討會）。我發現這個部分的討論資源蠻多的，所以最終把兩個部分的議題加以相連。人工智慧與法學，這部分的討論連結到的是基礎

doi:10.6342/NTU202103631

法學中很根本的問題，就是「到底什麼是法律」、「法律適用在做什麼」。在這個部分的討論中，我補了在政大時完全沒修過的法理學，以及挖了一下現在臺灣已經沒什麼人在討論的法律論證理論，甚至進入了法律解釋學的部分。雖然這個部分我沒有什麼貢獻，但我覺得在閱讀的過程中，自己有蠻多的回饋以及反思的，比較知道法律適用在做什麼事情（考上國考其實只要有類似人工智慧的模仿能力就差不多了）。

***

學無止盡，期許自己能夠繼續努力。

呂胤慶

2021 年 10 月 26 日

於宏德法律事務所

doi:10.6342/NTU202103631

doi:10.6342/NTU202103631

摘要

第二波人工智慧的廣泛應用，激起了法律與科技之間的緊張關係。人工智慧的建置目的，在於輔助人類更有效率從事任務，甚至根本的取代人類。然而，法律適用的任務是否能由人工智慧加以取代，有所疑義。本文的問題意識在於釐清，透過人工智慧從事法律適用的任務，是否與憲法的價值有所違背。關於這個問題意識，歐盟的一般資料保護規則已經作成了初步的價值判斷，原則上禁止以人工智慧從事法律適用的任務。這個價值判斷是否正確則必須進一步釐清。基於此，本文的研究目的即在探究憲法法治原則關於法律適用任務的要求，藉以評價人工智慧的法律適用，並審視歐盟立法者的前述立場，最後再針對「公部門使用人工智慧適用法律是否牴觸憲法」的問題提出回答。

本文認為若使用人工智慧從事法律適用的任務，將造成無法對於新的個案從事法律適用，以及區分個案的差異進行法律的續造。這樣的結果與憲法法治原則有所違背。換言之，歐盟法對於人工智慧的管制具有正當性。基於此立場，本文指出，公部門僅有在「人為介入」的條件中，方能夠利用人工智慧輔助公部門從事法律適用的任務。為確保人為介入的目的能夠達成，本文主張應揭露人工智慧在建置階段中的重要資訊，使公部門中的決定者能在個案中實質的評估是否採納或拒絕人工智慧的輔助。

在論證順序上，本文於第二章說明「人為介入」的內涵。人為介入目的在於分配機器與人類在任務執行上的互動關係，以人類作為任務課責的對象。在規範模式上，人為介入可以分為「機器輔助人類作成決定」與「人類事後推翻機器決定」的兩種類型。以目前歐盟一般資料保護規則中的「受人為決定」為例，本文指出「人為介入」管制模式背後的事實判斷是「人為決定」與「人工智慧的決定」有所不同，人工智慧無法完全取代人為決定。換言之，人為介入的目的在於維持現狀，維持以「人類」作為最終決策者的決定方式。

為釐清蘊含於「人為介入」的事實判斷是否正確，本文於第三章分析人為與人工智慧在適用法律、作成決定的任務上是否有差距。在釐清人工智慧的本質，及其在運作上的特色後，本文以法律適用的內涵作為基礎，指出人

doi:10.6342/NTU202103631

工智慧與人類決定的差異。針對人工智慧在運作上的特性，本文指出人工智慧在從事法律適用任務上所生的兩個問題：一、沒有辦法針對新個案從事法律適用；二、沒有辦法區分個案之間的差異從事法律之續造。

在說明人工智慧對於適用法律過程所造成的影響後，本文於第四章從法治原則出發，從規範面上評價這些特色。本文檢視法律續造在法治原則中的地位，並分析公部門若使用人工智慧從事法律適用的任務，是否存在不用進行法律續造的特殊原因。在法治原則在現今的社會中仍有其重要性的前提下，本文認為「人為介入」的規範性要求，使適用法律、作成決定的過程，能夠由人類區分事實的差異而作成個案的決定，保持規範適用典範變遷的可能性，並進一步證成人為介入是公部門正當使用人工智慧的必要條件。為確保「人為介入」在現實中能夠有效的落實，本文主張應建構人工智慧設計階段為揭露對象的透明法制框架，透過揭露人工智慧在建置階段之中的重要資訊，使人類決策者能夠有效的評估個案之中是否採納或者拒絕人工智慧的決定，以有效的落實人為介入的規範性要求。

關鍵字：人工智慧、自動化決定、演算法、巨量資料、公部門、人為介入、歐盟一般資料保護規則、法治原則、法律的續造、原則與規則、課責、透明、說明權、具說明能力的人工智慧

doi:10.6342/NTU202103631

ABSTRACT

A decade ago, nobody could have envisaged such a world, in which robots replaced the majority of laborers. Now, the widespread usages of data-based Artificial Intelligence (AI), is already intensifying the tension between law and technology. One of the harshest issues is whether AI can replace humans in the field of legal reasoning. In this thesis, the author will explore the differences between AI and humans when they conduct legal reasoning, and analyze under what conditions the public sector can use artificial intelligence legitimately. The author indicates two failures AI makes in legal reasoning. First, AI cannot apply the law on the new facts that have never taken place before. Second, Al is incapable of distinguishing between cases and creating new reasons to foster the change of the law. With these two failures, the thesis argues that the public sector violates the constitutional principle of the "rule of law" when using AI to apply the law automatically. Finally, the thesis tries to construct a regulatory framework including the obligations of human intervention and disclosure of formational information of AI. Under this human-centered framework, the public sector can enjoy the benefits brought by AI and, simultaneously, conform to the obligation set forth by the rule of law.

Chapter 2 illustrates what "human intervention" means. Among all the dazzling AI regulation proposals, one of the most familiar approaches is keeping a human in the loop of AI. Unfortunately, circling in the fog for the right of explanation, scholars across the Atlantic shed little light on the basic meaning of the human intervention regulatory model in the European Union General Data Protection Regulation (GDPR). The thesis contributes to the scholarship with a thorough analysis of this model. Interestingly, it shows that underlying this model is the implicit legislative intent that human-made decisions and AI-made decisions are totally different. Moreover, instead of following the global trend of AI fanaticism, the legislator decides not to jump on the train of technology revolution, but embrace the classic decision approach, human-made decision. This regulatory measure, obviously, raises a deeper question: what are the differences between human decisions and AI decisions?

Chapter 3 discusses the question posed by the EU legislator directly. Under practical scenarios, AI has two main aims. One is technological: imitating what humans can do. The other is scientific: answering what humans don't know. The author distinguishes between AI and humans on the work of legal reasoning, with a focus on AI with imitating ability. A "human" jurist in legal reasoning should give a word-elaborated reason on what

doi:10.6342/NTU202103631

law shall be interpreted under certain situations, justify a meaning-based extension of the legal requirements to be applied under certain facts, and vindicate a principle-based requirement of the legal rules that shall be recognized under a certain legal order. In other words, the law, as an institution, possesses the ability to adapt along with the variances between facts and the change of society. Contrasting these characteristics with AI, the imitating AI cannot in theory apply new facts, distinguish interpretations between individual cases, and contribute to changing the legal paradigm continually. However, what these distinctions shall mean under the constitution becomes a significant question.

Chapter 4 reviews these distinctions under the "rule of law". Based on the preconditions that the law includes both rule and principle, legal practitioners, along with judges and executive body, have the obligation to apply the law according to the facts and to attune the interpretation in line with the environment. The public sector would be in violation of the rule of law if it uses AI to apply the law directly without justification, for AI is incapable of altering legal interpretation itself. This thesis does not call for a complete ban on AI decision-making, but to impose conditions on the use of AI. Considering the importance of humans' ability, the author argues that a "human intervention" model is a suitable safeguard for constructing an accountable and organic legal reasoning system. Maintaining the human subject as the decision-maker and allowing the assistance of AI are the core obligations of this model. The issue this model triggers, nevertheless, is how to fulfill good human-computer interaction. The key to ideal human-computer interaction, in my opinion, is transparency. Seeing with knowing. The disclosure of the formational information of AI, such as information on data collection (e.g. collection methods, sample range, labelling criteria, set-cleaning methods) and algorithm (e.g. reason for choosing a certain algorithm), helps the decision-maker evaluate whether to accept AI assistance by examining the relationship between every individual case and the database.

Chapter 5, finally concludes with some key points extracted from each chapter, clarifies the limits of this study, and sets future research agenda.

Keywords: Artificial Intelligence; automated decision-making (ADM); algorithm; Big Data; the public sector; human intervention; Europe Union General Data Protection Regulation (GDPR); the Rule of Law; legal construction (Rechtsfortbildung); principle and rule; accountability; transparency; the right to explanation (RtE); explainable Artificial Intelligence (xAI)

doi:10.6342/NTU202103631

簡目

doi:10.6342/NTU202103631

doi:10.6342/NTU202103631

詳目

doi:10.6342/NTU202103631

doi:10.6342/NTU202103631

第一章

緒論

【案例：智慧國家案[1]】

若政府立法容許使用自動化決定來從事國家任務。司法機關與行政機關分別自行製作以機器學習為原理的自動化決定系統。

司法機關之中，法院主要推行智慧法官，由人工智慧進行個案的審判，認定個案事實是否符合法律要件，並作成是否成罪的判決。事實的認定與法律解釋的過程之中皆無人為的介入。司法部門所使用的自動化決定系統，主要依據全國法官過去對於同一條所作成的所有判決作為訓練資料，建置出能夠審判的全自動化決定系統。

在行政部門之中，金融監督管理委員會（下稱金管會），使用自動化決定，用於發現證券市場中影響交易秩序的非法行為[2]。金管會透過觀察過去的交易行為資料，藉以分析出其他可能與影響交易秩序行為有關聯性的交易行為模式。在運用上，金管會也是採取全自動化決定的方式，由系統自行判斷是否該當危害秩序的要件，直接作成行政處分限制交易者進行交易。

---

[1] 在美國，關於公部門使用自動化科技的討論中，自2005年即有相關文獻探討電子化政府對於正當法律程序的影響，甚至有提出「科技上正當法律程序（technology due process）」的概念。相關討論的脈絡，請參見劉靜怡（2021），〈科技正當法律程序的憲法意涵—美國判決與學說發展的檢視〉，收錄於氏著：《網路時代的隱私保護困境》，頁356-411，臺北：元照。但就筆者的觀察，美國近期之所以大幅度的討論人工智慧是否能夠於公部門中加以使用，可以說是肇因於2016年美國威斯康辛州最高法院*State v. Loomis*的案件，司法於審判過程中利用自動化科技輔助作成個案決定，而引起憲法正當法律程序的相關討論。基於此原因，本文選擇以「司法」利用人工智慧作為討論案例。*State v Loomis*, 881 N.W.2d., 749 (Wis 2016). 關於本案件的分析，*see e.g.,* Comment, *State v. Loomis: Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing*, 130 HARV. L. REV. 1530, 1530-1537 (2017); Han-Wei Liu, Ching-Fu Lin & Yu-Jie Chen, *Beyond State v Loomis: Artificial Intelligence, Government Algorithmization and Accountability*, 27:2 INT'L JL & IT. 122, 126-133 (2019).

相較於司法運用人工智慧的目的在於追求穩定、不恣意的個案裁判，行政部門利用人工智慧的目的主要在於強化行政效能。基於此目的上的差距，行政機關通常會採取不同種類的人工智慧，以發現特定的行為事實為目的。其中美國證券交易委員會即廣泛的透過人工智慧，從眾多的個案中來辨認違法的證券交易行為，並加以管制。本案例設計即以我國的相應組織「金管會」作為案例。*See* Scott W. Bauguess, The Role of Big Data, Machine Learning, and AI in Assessing Risks: A Regulatory Perspective, U.S. SECURITIES AND EXCHANGE COMMISSION (Jun. 21, 2017), ogy.de/ixa1 (last visited: Oct. 12, 2021).

[2] 證券交易法第148條：「於證券交易所上市有價證券之公司，有違反本法或依本法發布之命令時，主管機關為保護公益或投資人利益，得命令該證券交易所停止該有價證券之買賣或終止上市。」

doi:10.6342/NTU202103631

問題一：法院與金管會使用自動化決定作成個案決定，與憲法的法治原則是否有違背？

問題二：若前述問題的答案為是（違憲），法院與金管會在何種條件下使用自動化決定方屬合憲？

doi:10.6342/NTU202103631

1.1　問題意識與研究目的

資料（data）以及演算法（algorithm）正在形塑社會的樣貌。

過去，人類社會中的各種決定都是由人類自己作成，透過人類的眼睛觀察事情、透過人類的腦袋思考問題、透過人類的身體加以形成。然而，此種理所當然的情景，逐漸發生改變。

拜資料儲存成本的降低、電腦計算速度的指數倍成長、網際網路的普遍之賜，技術的進展造就了能夠分析大量資料以及進行高維度資訊處理的「人工智慧（artificial intelligence）」，以機器學習，自主從資料之間學習關聯性，模擬人類的行為模式。[3]前述機器學習的「演算法」，在技術上的整體提升，使人工智慧能夠更客觀、更精確、更迅速的對於事實做出判斷與預測，且有能力取代普遍被認為受偏見影響、決定效率慢、背景事實解讀不全面的人為決定。

這種以機器學習為中心的演算法，以不同的形式運用於公部門與私部門之中：透過歷史犯罪資料判斷犯罪熱點，事先部署警備資源以預防犯罪[4]；在刑事司法程序中分析過往案例，以決定量刑輕重、羈押時間的長短[5]；在大型流行病的控制上，透過分析巨量數據判斷可能的傳染路徑[6]；在稅務行政中，透過自動化程序對於不同納稅人核課稅捐[7]；銀行業者透過自動化技術決定是否貸款與特定客戶[8]；企業運用自動化技術，自大量的履歷中挑選最符合企業需求的應徵者[9]；社群媒體透過蒐集用戶的瀏覽行為，並以自動化程序來決定

[3] 關於機器學習的內涵請參見1.3.3.1以下的介紹。

[4] Asha Barbaschow, NSW Police using artificial intelligence to analyze CCTV footage, ZD NET (Jun. 6, 2021), ogy.de/7bax (last visited: 1, Jul. 2021).

[5] Priya Dialani, AI Will have Robot Judges Soon. What about Human Judges?, ANALYTICS INSIGHT (Mar. 4, 2021), ogy.de/qkj2 (last visited: 1, Jul. 2021).

[6] Keith Darlington, How Artificial Intelligence Is Helping Prevent the Spread of the COVID-19 Pandemic. BBVA (May 22, 2020), ogy.de/8c1q (last visited: 1, Jul. 2021)

[7] Richard Rubin, AI Comes to the Tax Code, THE WALL STREET JOURNAL (Feb. 26, 2020), ogy.de/qqvq (last visited: 1, Jul. 2021).

[8] Shannen Balogh and Carter Johnson, AI can help reduce inequity in credit access, but banks will have to trade off fairness for accuracy — for now, Insider (Jun. 30, 2021), ogy.de/dw3q (last visited: 1, Jul. 2021).

[9] David Windley, Is AI The Answer To Recruiting Effectiveness?, FORBES (Jun. 16, 2021), ogy.de/72hy (last visited: 1, Jul. 2021).

doi:10.6342/NTU202103631

動態消息的構成，影響閱聽者所接收到的訊息[10]。以上人工智慧的運用案例呈現了一個發展上的趨勢：**人工智慧取代人類勢不可擋**。過去由人類從事的工作，逐漸可以委託給成本低、速度快的人工智慧加以執行。因此，國際間的公部門以及私部門，皆開始廣泛利用自動化技術，輔助相關任務的執行，而「**法律的適用**」也逐漸成為下一個能夠由人工智慧完全從事的任務。[11]

然而，國際學界間對於人工智慧應如何管制的討論[12]，除了減緩我們對於人工智慧的狂熱外，也道出了一個重要的事實：**以機器學習為技術內涵的人工智慧，並非不會出問題的完美決策工具**。人工智慧在應用上會出問題的原因，在於它是透過資料間關聯性的竅門，來面對以人類經驗為中心的類比社會。因此，在人類社會中該如何應用人工智慧，就應該以知悉「資料關聯性」的侷限為前提。唯有釐清人工智慧的侷限與特定任務之間的關係後，才能對於人工智慧是否應該能夠從事人類的任務，作出適切的評價。[13]換言之，除了

---

[10] Thomas Macaulay, Here's how AI determines what you see on the Facebook News Feed: Facebook shared new insights on how algorithms power the feed's ranking, NEURAL (Jan. 26, 2021), ogy.de/51qk (last visited: 1, Jul. 2021).

[11] *See* Benjamin Alarie, *The Path of the Law: Towards Legal Singularity*, 66 U. TORONTO L. J. 443 (2016).

[12] 例如：經濟合作暨發展組織（Organization for Cooperation and Development, OECD）於2019年提出《人工智慧原則》；歐盟則分別於2018至2021年間提出《歐盟人工智慧策略》、《值得信賴的人工智慧準則》、《人工智慧白皮書》以及統合各會員國關於人工智慧的規範草案，企圖制訂人工智慧在歐盟使用的一體性框架；美國聯邦政府除了於2021年正式成立國家人工智慧倡議辦公室著手擬定美國人工智慧的相關政策外，美國國會政府課責總署也正式提出人工智慧於聯邦政府各部門使用的課責框架；加拿大政府於2019年對於公部門的自動化決定使用則提出《自動化決定指令》規範公部門內部對於自動化決定的使用；英國資訊專員辦公室於2020年針對人工智慧的使用提出《人工智慧與資料保護指引》；新加坡的個人資料保護總署則於2019年提出模範人工智慧治理框架，針對私部門以及公部門人工智慧的使用要求提出框架性的說明，並於2020年加以修訂。以上的資料來源請參見第二章註腳7至註腳13。

[13] 本文認為法律應如何面對人工智慧，可能立基於兩種完全不同的問題意識，而這兩種問題意識分別會出現兩種不同的論證方式。第一種問題意識是立基於人工智慧與人類在能力的品質上**仍舊有差距**或者人類與人工智慧間的**差異具有重要性**出發，透過說明人工智慧與人類在決定品質上的差距，並進一步論證這種品質差距影響了特定任務的本質；目前大宗的文獻都是從此觀點出發，透過仔細的釐清人工智慧的能力上的不足，以證成特定任務不應該使用人工智慧。第二種問題意識則是立基於人工智慧在能力上已經與人類**沒有差距**，或人類與人工智慧之間的**差異已經不具有重要性**，此時涉及的問題不在特定差距對於任務的影響，而在於應該如何認知人工智慧與人之間的關係，狹義的來說就是「人機互動」的問題；目前已經開始有學者對於此部分有加以研究，其中直接以此種問題意識處理者，*see*, *e.g.*, Kiel Brennan-Marquez & Stepen E. Henderson, *Artificial Intelligence and Role-Reversible Judgement*, 109 J. CRIM. L. & CRIMINOLOGY 137 (2019).也明確指出此種問題意識區分者，*see*, *e.g.*, Lyria Bennett Moses, *Not a Single Singularity*, *in* IS LAW COMPUTABLE?: CRITICAL PERSPECTIVES ON LAW AND ARTIFICIAL INTELLIGENCE 213-220 (Simon Deakin & Christopher Markou eds., 2020). 當然，這兩種問題意識不一定能夠截然區分，尤其在某些「人機互動」的特色會直接影響任務的品質的情形。若要詳細的討論人工智慧與法律的關係，討論的方式應從第一個問題

doi:10.6342/NTU202103631

避免人工智慧從事任務所伴隨的傷害外，討論的焦點反而應該著重於：**習得資料間關聯性的機器學習，是否適合從事特定任務**。如果使用這種人工智慧根本無法達成特定任務的要求，但在規範上卻盡力想辦法避免人工智慧在執行特定任務中所造成的傷害時，反而是緣木求魚。因為有些任務可能根本不應該由人工智慧越俎代庖。

如果以公部門作為使用人工智慧的主體，關鍵性的問題尤其是公部門是否能夠使用人工智慧從事適用法律的任務，理由在於法律適用對於社會影響的無遠弗屆。因此，本文的問題意識即在於釐清，**透過人工智慧從事法律適用的任務，是否與憲法的價值有所違背**。

若以前述的問題意識為中心，對國內法學界以人工智慧為對象的相關研究進行回顧[14]，可以發現國內學者方面已經開始注意到這個問題的不同面向。

意識出發，分析在特定任務之中人工智慧與人類是否不同。若這個問題的答案為不同，則可進一步對於這個不同進行規範性分析（路徑一）；若這個答案為相同，則必須進一步從第二個問題意識加以分析（路徑二）。國內採取路徑二，並且論述得相當清楚的文獻，請參見Luís Greco（著）、鍾宏彬（譯）（2021），〈沒有法官責任的法官權力：為什麼不許有機器人法官〉，《月旦法學雜誌》，315期，頁170-195。

[14] 我國對於人工智慧所生的一般性法學議題盤點的文獻請參見參見例如：劉靜怡（2020），〈人工智慧時代的法學研究路徑初探〉，收於：李建良編，《法律思維與制度的智慧轉型》，頁91-134，臺北：元照；李建良（2020），〈人工智慧與法學變遷 - 法律人面對科技的反思〉，收於：李建良編，《法律思維與制度的智慧轉型》，頁1-87，臺北：元照；林勤富（2018），〈人工智慧法律議題初探〉，《月旦法學雜誌》，274期，頁195-215；劉靜怡（2018），〈人工智慧潛在倫理與法律議題鳥瞰與初步分析〉，收於：劉靜怡編，《人工智慧相關法律議題芻議》，頁1-45，臺北：元照。

若將我國關於人工智慧與法學的研究進行分類，除了一般性法學議題與人工智慧盤點的文獻，依據問題意識可以大別為三種類型的研究，分別是「人工智慧與法學理論（A）」、「人工智慧與法律制度（B）」以及「人工智慧與法律實踐（C）」。

「人工智慧與法學理論（A）」的相關研究主要在釐清人工智慧對於法學基礎的知識上有何影響，例如：人工智慧是否能夠從事法學推論、人工智慧從事法學推論會產生何種結果。參見例如：陳弘儒（2020），〈初探目的解釋在法律人工智慧之運用可能〉，《歐美研究》，50卷2期，頁293-347；陳弘儒（2020），〈初探目的解釋在法律人工智慧系統之運用可能〉，收於：李建良編，《法律思維與制度的思維轉型》，頁227-302，臺北：元照；邱文聰，前揭註29，頁137-168；黃銘傑（2019），〈人工智慧發展對法律及法律人的影響〉，《月旦法學教室》，200期，頁51-54。

我國大多數的法律文獻則著重於「人工智慧與法律制度（B）」的分析，以人工智慧為新興的案例，探討個別法領域中的制度內涵是否因為人工智慧有加以改變的必要，或更進一步的說明法律應該如何積極的面對人工智慧帶來的挑戰。近期由劉靜怡教授所編輯的《人工智慧相關法律議題芻議》一書，與李建良教授所編輯的《法律思維與制度的智慧轉型》一書，多為此種取徑。（劉靜怡教授所編書中採取此種取徑的學者諸如：邱文聰教授、吳從周教授、李榮耕教授、沈宗倫教授、顏厥安教授、黃居正教授；李建良教授所編書中採取此種取徑者諸如：吳全峰教授、何之行教授、廖貞、楊岳平教授、沈宗倫教授、林勤富教

doi:10.6342/NTU202103631

例如邱文聰教授在〈亦步亦趨的模仿還是超前部署的控制？AI 的兩種能力和它們帶來的挑戰〉一文中，已經注意到人工智慧在使用上產生的一般性問題。邱教授區辨出人工智慧「複製模仿」與「發掘資料間關聯性」兩種能力，並分別對於這兩種能力所造成的社會爭議加以說明。[15]邱教授認為人工智慧的「複製模仿」的能力，不僅過程無法理解，在特定任務之中（例如：文字意涵的釐清）有所侷限，甚至無法「打破成規」以促成創新、甚至是典範變遷。[16]相較於複製模仿，人工智慧的「發掘資料間關聯性」之所以在應用的面向會產生社會爭議，理由在於人工智慧所發掘的資料關聯性欠缺「外部的可解釋性」，造成人類為了滿足控制的慾望，過度肆無忌憚的使用這種無法確定內涵為何的關聯性。[17]

延續這樣的觀察，邱教授在〈第二波人工智慧知識學習與生產對法學的挑戰〉一文中，討論人工智慧對於法律適用任務的影響。邱教授認為具有模仿能力的「監督式學習」，以人類知識為基礎，雖然能夠高度的模仿人類行為，但無法擺脫知識系統中已經存在的偏誤，也無法推翻先例，創造新的法學知識；而透過發掘資料間關聯性能力的「非監督式學習」，雖然能夠找到人為無法理解的超知識關聯性，但在應用的過程中造成，管制者為了促成特定結果的發生，採取「結果論證」，介入影響與結果具有關聯性的變因以達到控制的目的。[18]

授、李怡俐教授、王怡蘋教授與陳柏良教授。詳細書目內容請參照參考文獻）。

「人工智慧與法律實踐（C）」的研究則是從技術的面向出發，由下而上的分析人工智慧在何種程度上能夠對於法律的實踐有所貢獻 。這個部分的研究，是透過人工智慧發現法律人無法發現的趨勢，釐清人工智慧在多大的程度內能夠促進法律制度的發展。參見例如：參見例如：王紀軒（2019），〈人工智慧於司法實務的應用〉，《月旦法學雜誌》，293期，頁93-114；黃詩淳、邵軒磊（2017），〈運用機器學習預測法院裁判─法資訊學之實踐〉，《月旦法學雜誌》，270期，頁86-96；黃詩淳與邵軒磊（2019），〈人工智慧與法律資料分析之方法與應用：以單獨親權酌定裁判的預測模型為例 〉，《臺大法學論叢》，48卷4期，頁2023-2073；黃詩淳、邵軒磊（2020），〈以人工智慧讀取親權酌定裁判文本：自然語言與文字探勘之實踐 〉，《臺大法學論叢》，49卷1期，頁195-224。

[15]邱文聰（2021），〈亦步亦趨的模仿還是超前部署的控制？AI的兩種能力和它們帶來的挑戰〉，收錄於《智慧新世界？人文社科學人眼中的AI》，作者手稿，頁2（即將出版）。

[16] 邱文聰，前揭註15，作者手稿，頁4-5。

[17] 邱文聰，前揭註15，作者手稿，頁6-7。

[18] 邱文聰（2020），〈第二波人工智慧知識學習與生產對法學的挑戰—資訊、科技與社會研究及法學的對話〉，收於：李建良編，《法律思維與制度的智慧轉型》，頁156-163，臺北：元照

doi:10.6342/NTU202103631

若就這兩篇文章而言，邱教授是從人工智慧與人類之間有所差異的觀點出發，釐清人工智慧與人類在「法律適用」上所產生的差異。邱教授的觀察固然正確，但對於這兩種結果，在規範上應如何評價，尤其是法學中是否容許這些差異此（即法律適用是否根本不適合由人工智慧來加以處理的問題），是邱教授並未處理的部分。而這也是本文能夠繼續加以釐清的部分。

除了以上描述性的說明外，也有學者針對利用人工智慧取代人類是否對於憲法保障的價值造成影響進行規範性的評價。例如劉靜怡教授的〈人工智慧潛在倫理與法律議題鳥瞰與初步分析—從責任分配到市場競爭〉一文中，劉教授指出自動化決定在行政、刑事等國家決定的行使中，基於演算法的黑箱性質產生了責任分配的困難，應落實「正當法律程序」的要求，釐清「人」與「人工智慧」在決定中的責任分配關係。就此，劉教授認為「獲得解釋的權利（right to explanation）[19]」透過要求人工智慧的決定發揮「論證和說明決定理由」的功能，是落實正當法律程序價值的重要權利。[20]此外，邱文聰教授的〈初探人工智慧中的個資保護發展趨勢與潛在的反歧視難題〉一文，也是從憲法的觀點出發，評價人工智慧對於既有法律體系的衝擊，以及分析歐盟一般資料保護規則中的人為介入（human intervention）並要求提供說明（right to explanation），在規範面上是否是一個有效回應對於憲法挑戰的手段。[21]

劉教授與邱教授這兩篇在憲法脈絡上討論的文章，則可以看出另一個針對人工智慧與法律間關係的討論脈絡。這個脈絡主要以人工智慧取代人類從事任務為基礎，釐清人工智慧對於憲法所造成的影響，並提出符合憲法要求的人工智慧管制架構。這一類脈絡的討論中，通常會有一些沒有明確指出的預設。[22]第一個預設是以「國家」是使用人工智慧的對象為討論基礎。基於憲

[19] 對於「right to explanation」，目前中文法學界皆譯為「解釋權」或者「獲得解釋的權利」，指的是自動化決定形成理由的闡明。但「解釋」一語在法律界具有特定的意涵，尤其指涉的是對於規範意涵釐清的語意行為，與自動化決定下的explanation有所不同。因此本文在論述時提及自動化決定中的right to explanation時將以「說明權」稱之。此處由於是對於國內學者的文獻回顧，因此保留作者的原文。

[20] 劉靜怡，〈人工智慧潛在倫理與法律議題鳥瞰與初步分析〉，前揭註14，頁31-33。

[21] 邱文聰（2018），〈初探人工智慧中的個資保護發展趨勢與潛在的反歧視難題〉，收於：劉靜怡（編），《人工智慧相關法律議題芻議》，頁145-175，臺北：元照。

[22] 與其他文章相比，兩位教授在以下兩個論證上的預設，就講得非常清楚。

doi:10.6342/NTU202103631

法基本權利的「針對國家性」，若以憲法基本權利的保障針對人工智慧的使用進行評價，此時規範的評價對象僅限於「國家」使用人工智慧的情形。第二個預設是人工智慧造成的影響，是「以人類作為比較基礎」所生的影響。許多文章在討論人工智慧所造成的影響時，直接以人工智慧可能會造成何種基本權利的侵害作為釐清的重點。這樣的討論其實沒有釐清「專屬」於人工智慧的問題，因為人類決定也會造成基本權利的侵害。所以在探討人工智慧所造成的影響時，所謂的影響是與人類相比所生。

總結以上學者的論述，可以歸納出兩條主要的討論脈絡。第一個討論脈絡是描述性的，說明人工智慧對於法律適用任務本身的影響。第二個脈絡則規範性的，以憲法對於人工智慧的使用進行評價，以及試圖提出管制手段。這兩個脈絡尤其不是獨立的問題，因為精準的評價必須以正確的描述為前提。

如果要為這兩個脈絡建立一個合理的討論方向，本文認為以「國家」在進行「法律適用任務」的觀點出發，則可以將這兩個脈絡的討論緊密的相連，建構出一個相當清晰的討論脈絡。透過描述性的釐清人工智慧對於法學的改變與挑戰後，更進一步的從憲法的框架評價這種國家透過人工智慧，以進行法律適用，所產生的改變以及挑戰是否被容許加以說明。換言之，以上文獻回顧所凸顯的問題意識尤其是：**透過人工智慧從事法律適用的任務，是否與憲法的價值有所違背？**

關於此問題，歐盟的一般資料保護規則已經作成了初步的價值判斷。歐盟資料保護基本規則第 22 條[23]，賦予資料主體一個權利，能夠拒絕自動化決定所產生的法律效果。從本文的問題意識來看這個規範，可以得出，歐盟的立法者同時作成了不得以人工智慧從事法律適用任務的價值判斷。這個價值判斷是否正確則必須進一步釐清。

基於以上的問題意識，本文的**研究目的**即在探究憲法法治原則關於法律適用任務的要求，藉以評價人工智慧的法律適用，並審視歐盟立法者的前述立場，最後再針對「公部門使用人工智慧適用法律是否牴觸憲法」的問題提出回答。

[23] 關於本條的內容，請參見本文第二章的說明。

doi:10.6342/NTU202103631

以上的研究目的，可預期能有以下兩個貢獻：一、規範性的評價人工智慧對於法律適用內涵的挑戰；二、為公部門使用人工智慧提出符合憲法價值的管制手段，在享受人工智慧的利益之餘，符合憲法的要求。

在研究方法上，本文採取文獻分析法，就國內外學者針對相關問題的研究，進行分析以及檢討，並且進一步提出本文的論點。此外，由於本文最終將提出具體的管制手段，因此本文將同時採取比較法研究，分析外國針對人工智慧的法制框架優劣，釐清這些外國法制是否足供我國加以借鏡。

doi:10.6342/NTU202103631

1.2 研究範圍

在以上問題意識、研究目的的提出後，此處可以將本文的研究範圍做更明確的界定如下：

1.2.1 「公部門」的決定

第一，本文所討論的人工智慧使用者以公部門為限。由於公部門的任務推行具有強制力及執行力，能夠影響整個社會，因此本文將限縮討論的對象於公部門的人工智慧運用，以拘束公部門的憲法框架作為判斷是否容許的基準。

雖然人工智慧的特色，並不會因為公部門與私部門的不同而有所差異。但對於私部門運用人工智慧的管制，其管制基礎何在？為何私部門運用自動化決定應加以管制？這在討論方法上與公部門的討論有所不同。[24]雖然私部門的重要性以及影響力在近年來逐漸提升，憲法對於基本權的對抗對象是否應及於特定強力的私部門在近來的各種議題之中討論熱烈，但為了集中本論文討論的內容，本文不會處理私部門應用人工智慧所產生的相關管制爭議。

1.2.2 適用法律產生效果的決定

第二，本文所探討的人工智慧應用任務**以適用法律，並且對外直接產生影響人民權益法律效果的決定**為限。

政府運用人工智慧從事任務的類型多端，除了「適用法律作成決定」的任務之外，從內部的政策形成到外部的決定作成，過程包括了對於事實的認定、以及法律評價以及手段的選擇、單純資訊性的提供、軟性管制手段的採取、資源的部署…等其他不涉及直接適用法律的任務。基於適用法律、作成決定任務對於人民所造成影響的「直接性」，因此本文討論的對象限縮於這些利用人工智慧，輔助進行適用法律所作成的決定。比較接近本文討論的決

[24] *See*, *e.g.*, Sonia K. Katyal, *Private Accountability in the Age of Artificial Intelligence*, 66 UCLA L. REV. 54, 54-141 (2019).

doi:10.6342/NTU202103631

定類型，諸如：行政機關所作成的行政處分、司法機關所作成的裁判，等其它國家針對人民直接作成的決定類型。

其他並非涉及適用法律作成決定的政府行為，例如抽象的國家行為（法律、行政規則的制定），或其它不具有法律效果但具有實質上影響力的事實行為，雖然並非本文直接回應的對象，但憲法的保障是無漏洞的。因此利用人工智慧輔助進行這些任務時，仍必須符合憲法的這些要求。只是究竟應該如何使憲法落實到這些領域之中，則是必須另外處理的問題。

1.2.3　發現並應用「資料關聯性」爲原理的自動化決定

第三，本文討論的人工智慧技術內涵，限縮於**運用特定的演算法分析資料間關聯性**，且主要以**自動化決定**為形式的應用類型。

從字面上理解「自動化決定」的內涵，似乎只要決定的形成過程由機器為之，就相當於自動化決定，例如：街道上的紅綠燈、交通違規檢舉用的照相機。自動化決定在文義上當然包括前述這些應用。但本文中對於自動化決定應用的類型做了一個限縮。本文的自動化決定指的尤其是為了**達成人為指定的決定目的，透過演算法分析巨量資料，應用其分析結果（模型）的決定系統，且決定的過程之中並無人為的介入**。因此本文要討論的「自動化決定」，具有以下三個特徵：一、以巨量資料為基礎；二、透過演算法分析資料之間的關係；三、無人為的實質介入。

實際來說，「自動化決定」的用語並不是表彰特定一種技術，相反的他是一個法律術語。例如歐盟的一般資料保護規則法就以「自動化決定[25]」的用語作為規範的標的。這個用語的特色其實在說明了特定決定系統中無人為介入的事實[26]。由於本文的問題意識在於釐清「人為決定」以及「人工智慧作成

[25] 例如歐盟一般資料保護規則的第13條第2項第f款、第14條第2項第g款、第15條第1項第h款、第22條第1項的規定即以「自動化決定（automated decision-making）」作為法規之用語。關於以上這幾個規範內涵的介紹，請參見第二章的說明。

[26] 資訊學界會區分「automated」、「autonomous」以及「algorithmic」三個用語：automated所指涉者為無人為介入，使程式自行運作的情形；autonomous則是有人為操控機器；而algorithmic則是比較大的概念，只要決定中有使用到程式即屬之。*See* Maja Brkan & Grégory Bonnet, *Legal and Technical Feasibility of the GDPR's Quest for Explanation of Algorithmic Decisions: of Black Boxes, White Boxes and Fata Morganas*, 11 EUR. J. RISK REGUL. 18, 23-24 (2020).

doi:10.6342/NTU202103631

決定」的差異，因此「自動化決定」一語尤其描述的是純粹由人工智慧做成的決定，而無人為介入、或者是人類做成決定的情形。

此處衍生的問題是，何謂「人為介入」？現行仰賴人類建立的自動化決定系統之中，包括為數不少的人為決定，例如：演算法的類型採取、資料的選擇與標籤、系統的調校。本文所謂的「無人為介入」指的是，已經完成訓練、正在使用的自動化決定系統，其對於新的案例事實產出結果至結果現實中發生法律效果加以應用的過程中，人類決定者並沒有權利或義務實質審視自動化系統運算結果妥適與否，直接將自動化系統的結果充當於直接發生現實中法律效果的國家決定的情形，而非指涉前述自動化決定系統建立階段中，人類工程師所作成的大大小小的決定。

綜上所述，本文透過「自動化決定」這個對象進行範圍的限定，限縮討論標的在於純粹利用資料關聯性的決定方式與人類決定有何種差異。

doi:10.6342/NTU202103631

1.3　名詞定義

作為一篇討論「技術與法律（technology and law）[27]」關係的論文，若要精確的對於科技的使用進行法律評價，免不了必須更進一步的說明科技的內涵。此處要指出的是，以下對於「人工智慧」內涵的說明，將不會過度的集中在此種技術如何形成的，而會將重心放在這些技術在應用上會產生何種結果。

人工智慧一語泛指一切以「人類思想」為模擬對象的各種科技嘗試，因此其並非指涉特定的科技，僅為科技設計的特定目標。[28]以發展人工智慧的方式以及時序作為區分標準，在演進上可以區分為三個時期[29]：第一波的人工智慧所指涉的是七零年代到八零年代中期間以規則為基礎的人工智慧；第二波的人工智慧則是八零年代至今以資料為基礎的人工智慧；而第三波人工智慧則以發展具有獨立思考能力的「通用人工智慧（Artificial General Intelligence, AGI）」作為目標。第一波人工智慧因為大量知識形式化的困難以及面對現實多變的僵固，已經逐漸脫離學界的討論對象；而第三波人工智慧目前僅止於討論的階段，在實踐上尚無較突破的斬獲。在此發展脈絡之下，本文討論的對象即是著聚焦於第二波人工智慧所發生的應用爭議。以下簡單的說明第一波人工智慧到第二波人工智慧的發展，以及第二波人工智慧在應用會產生的結果。

話說從頭，資訊科學界一直以來致力研究如何透過電腦取代人腦。第一波人工智慧的發展，是以「知識」為系統核心的運算結構，將特定領域的知

---

[27] 「Technology」一語在臺灣中文語境下經常以「科技」稱之，然而這個翻譯是否適切可以再斟酌。理由在於中文的「科技」一詞其實包含「科學」與「技術」兩個意涵。而在英文語境下的科技指的尤其是「Science and Technology」。換言之，technology其實比較偏向「技術」的意思。這個字眼的誤用其實會造成將技術作為科學的結果，混淆「技術物」與「知識」兩者之間的關係。因此，本文在使用technology一語時，盡量皆以「技術」稱之。

[28] 人工智慧的概念可以溯源至1950年Allen Turing所提出的「圖靈測試（Turing Test）」以解決機器是否能思考、具有智慧的問題。而「人工智慧（Artificial Intelligence）」一語，則是由John McCarthy於1956年8月的Dartmouth會議中加以提出。其將人工智慧定義為「用以製造出智慧的科學技術（the science and engineering of making intelligent machine）」，*see* Stan Franklin, *History, motivation, and core themes*, *in* THE CAMBRIDGE HANDBOOK OF ARTIFITIAL INTELLIGENCE 18 (Keith Frankish & Willaim M. Ramsey eds., 2014).

[29] 中文法學文獻對於人工智慧的演進可參見邱文聰，前揭註18，頁139-143；John Launchbury, A DARPA Perspective on Artificial Intelligence, ogy.de/nh5g (last visited: Jul. 16, 2021).

doi:10.6342/NTU202103631

識轉換成數學公式，並置入系統之中。[30]使用者將個案置於「知識為基礎（knowledge-based）」（或稱為以「規則為基礎（rule-based）」）的演算法中加速運算的過程而得出結果。此種模型在發展上，雖然有效的提升運算的速度，但人類知識的浩瀚以及多變造成此種演算法在使用上的侷限。

伴隨著網際網路的出現、個人資料取得的困難性降低，電腦資訊科學界即開始發展以「資料為基礎（data-based）」或「資料驅動（data-driven）」的第二波人工智慧。而近期法政、社會學界熱烈討論的「演算法[31]」，即為此種類型的人工智慧。就實際而言，第二波人工智慧下的演算法侷限找出資料與資料之間關聯性（correlation）或模式（pattern）的類型[32]。透過演算法，分析資料與資料之間人類無法察覺的關聯性，並進一步的加以應用。蘊含於這種應用方式的理念是，事實之間存在一定的關聯性以及規律，而這些規律會重複的出現；若利用這些規律，即可用於未來事件的判斷以及預測。這種要求機器「主動學習」、「提取」資料之間關聯性的方式，被稱為「機器學習」。[33]相較於「以規則為基礎」的第一波人工智慧是將資料輸入已經知悉的公式之中獲得結果，以機器學習為基礎的第二波人工智慧，則是自過往的案例之中提取出公式，用以應付未來的事實。

在應用上，目前的垃圾信件辨識、銀行放貸與否、語音辨識、圖像辨識等技術，大多是以機器學習作為應用方式。而法學界目前所討論的演算法、人工智慧管制爭議，例如人工智慧可能產生的歧視現象，即是聚焦於辨認資料間關聯性的機器學習所生。[34]

---

[30] *See* Reuben Binns, *Algorithmic Accountability and Public Reason*, 31 PHIL. & TECH. 543, 545 (2017).

[31] 現今法學界討論的演算法與演算法在資訊科學界原始的意義已經有些微的偏離。在資訊學界上，所謂的演算法，其內涵非常單純，係指：藉由輸入特定的要求（input），而能夠得出一定的結果（output），以解決所設定任務的數學公式，而不限定任務為何。*See* ETHEM ALPAYDIN, INTRODUCTION TO MACHINE LEARNING 2 (3 ed. 2014). 所以本文認為以演算法稱呼人工智慧是一個不太精準的描述。

[32] *See* Karen Yeung, *Algorithmic regulation: A critical interrogation*, 12 REGUL. GOV. 505, 505 (2018).

[33] *See* ALPAYDIN, *supra* note 31, at 2.

[34] *See*, *e.g.*, Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671 (2016); Anupam Chander, *The Racist Algorithm* 115 MICH. L. REV. 1023 (2016); Talia B. Gillis & Jann L. Spiess, *Big Data and Discrimination*, 86 U. CHI. L. REV. 459 (2019); Pauline T. Kim, *Data-Driven Discrimination at Work*, 58 WM. & MARY L. REV. 857 (2017); Sandra G. Mayson, *Bias In, Bias Out*, 128 YALE L. J. 2218 (2019); Anya E.R. Prince & Daniel Schwarcz, *Proxy Discrimination in the Age of Artificial Intelligence and Big Data*, 105 IOWA L. REV. 1257 (2020); Aziz Z. Huq, *Racial Equity in Algorithmic Criminal Justice*, 68 DUKE L.J. 1043 (2018).

doi:10.6342/NTU202103631

第二波人工智慧雖然在理論以及現實面上的討論聲勢如日中天，但其在「智慧」的內涵上仍具有本質的缺陷：無法建立因果關係（causality）。因果關係的建立，必須透過持續的觀察，發現規律改變的因素；透過主動介入調整因素的發生製造預期的結果，最終透過想像的能力來思索若特定因素沒有調整對於預期結果發生何種影響。[35]以資料關聯性為基礎的第二波人工智慧，目前僅能透過持續的觀察發現可能促成規律改變的因素，而沒有辦法介入運作過程中驗證或透過思索建立起因果理論；因而，第二波人工智慧也沒有辦法達成人類智慧的程度。不過第二波人工智慧的本質缺陷，並不妨礙本論文的論述，因為本文的問題意識在於釐清傳統以人類智慧為重心的公部門，是否有可能接納第二波人工智慧作為任務執行的工具。

若要對於第二波人工智慧進行更詳細的分析，首先要釐清機器學習的內涵。在分類上，依據學習的方式，主要可分為以下三種類型：分析人為事先定義（label）資料所代表意涵並自行學習資料特徵（feature）與意涵關係之監督式學習（supervised learning）、分析資料特徵之間關係之非監督式學習（unsupervised learning）以及讓機器與環境自行互動並透過回饋的方式進行試錯學習，從互動過程中透過取得回饋的強化式學習（reinforcement learning）。[36]

在應用方式上，**監督式學習**功能主要在習得事件的規律，並將此規律應用於新的情況中作成預測（進行分類），應用領域包括市場中的股價或房價預估、語音辨認、垃圾郵件…等其它能形塑為分類的任務。[37]監督式學習決定較精準，但仰賴大量事前人工對於資料的處理。相較於監督式學習的目的在於釐清不同變項對於結果的影響程度高低，**非監督式學習**功能在於釐清不同變項之間的關係高低，為資料群與資料群之間的關係作成區分，判定特定資料屬於 A 資料群或 B 資料群，應用領域包括購物網路上推薦購買商品（特定消費模式屬於 A 類型消費者或 B 類型消費者）、違章行為的發現（特定行為

[35] *See* generally JUDEA PEARL & DANA MACKENZIE, THE BOOK OF WHY: THE NEW SCIENCE OF CAUSE AND EFFECT 2018).

[36] *See* CHRISTOPHER M. BISHOP, PATTERN RECOGNITION AND MACHINE LEARNING 2-3(1 ed. 2006).

[37] *See* Jamie Berryhill, et al., *Hello, World: Artificial Intelligence and its Use in the Public Sector* 48-51 (OECD Working Papers on Public Governance), available at: ogy.de/v9ie.

doi:10.6342/NTU202103631

模式屬於有從事違章行為者或無從事違章行為者）無從事違章行為者。[38]非監督式學習的優點在於無庸事前對於資料進行人為的標記，並且可以察覺人類無法辨認的關聯性；但其缺點在於若無足夠大的訓練資料時將不夠精準，且特徵若增加，資料群與資料群的關係將變的相當複雜。透過與環境進行試錯學習的**強化式學習**，能夠從事複雜的任務，然而其仰賴更大量的訓練資料，相關的應用例如自駕車的開發、Alpha go 以及機器人的製作。[39]

以上關於機器學習的分類方式，主要是基於應用「技術」差異所生。縱然這樣的分類方式，可以讓不具有相關知識的人，很快速的掌握不同領域中所應用的機器學習技術類型；但這種分類方式，因為並未著眼於特定學習方式在社會中的意義，因此這組技術上的分類，對於法學分析的目的幫助有限。

相較於這種技術的分類，邱文聰教授對於機器學習在社會中的能力，有進一步的觀察以及分類。邱教授對於機器學習的應用歸納出了兩種核心能力，這兩種能力分別是：「複製模仿（emulation）」與「發掘資料間關聯性（association discovery）」。[40]複製模仿能力指的是機器學習以「聯結主義（connectionism）」為原理，從大量既存的先前案例中，找出且複製這些案例之中的聯結關係，達到模仿人類行為目標的運作方式。[41]這種能力的運用，範圍從人類身體動作、認知思維、知識系統甚至到決策行動，都有應用的空間，增加相關運用的效率。相較於複製模仿能力，機器學習的另外一個核心能力是「發掘資料間關聯性」，透過重新建立資料的分類（clustering），在社會中局部的拓展人類既有知識的邊界。[42]在應用上，發掘資料間關聯性的能力，能夠用於「預測」的任務之中，在因果關係較遠或者不清楚有無因果關係的兩事件之間，建立關聯性，以供社會各種任務的需求。[43]

邱教授的觀察以及分類，將機器學習在社會中的貢獻以及能力做了精確的連結。本文以下的論述，將立基於這個分類，進一步開展，進行法律分析。

---

[38] *Id*. at 51-53.
[39] *Id*.
[40] 邱文聰，前揭註23，頁2。
[41] 邱文聰，前揭註15，頁3。
[42] 邱文聰，前揭註15，頁5-6。
[43] 邱文聰，前揭註15，頁6。

doi:10.6342/NTU202103631

精確的來說，本文尤其將聚焦於具有模仿能力的人工智慧，釐清是否能夠透過此種方式從事適用法律的任務。依據本文前述對於「自動化決定」的定義，「以巨量資料為基礎」、「透過演算法分析資料之間的關係」、「無人為實質的介入」的三個特徵，能夠排除了人類的影響，進一步的聚焦於因為使用機器學習所產生的結果。

doi:10.6342/NTU202103631

1.4 本文論點

本論文的論點如下：**利用人工智慧從事法律適用的任務，將無法對於「新事實」從事法律適用，且無法因應個案在不同規範背景條件下的差異，從事法律的續造。因此，公部門若以人工智慧的方式從事法律適用的任務，將與憲法的法治原則有所牴觸。**

以上的論點並非完全的將人工智慧禁絕於公部門適用法律、作成決定任務外，而是為人工智慧在前述任務的應用中加上了若干的限制。若將以上的論點具體落實於**管制手段**，本文認為僅有在「**人為介入**」的條件下，公部門使用人工智慧從事適用法律的任務時，方符合憲法法治原則的要求。所謂的「人為介入」，指的是維持公部門在適用法律的任務時，「人類」作為法律適用的主體，人工智慧只能是輔助的角色，或在容許使用人工智慧從事適用法律的情況中，確保人類能夠介入人工智慧的運作過程中，推翻人工智慧的決定。人為介入作為一種管制的措施，除了維持了人類作為決策者的現狀外，同時也要求人工智慧在公部門的決策過程之中，僅能以輔助、備位的地位加以存在。此外，為了確保人為介入的目的能夠實現，本文主張必須建立以「**設計階段的重要資訊**」為揭露對象的透明法制框架，使（人類）決策者能夠透過這些資訊判斷個案中是否接受人工智慧的輔助。

此處必須指出的是，本文的論點、評價結果、管制主張是從理論的觀點出發，抽象的討論機器學習的能力發揮到「極致」的時候，法學應如何回應的論述。[44]

但本文的研究問題尚無法全面性的處理人工智慧的法律問題。本文認為，若要完整的釐清機器學習在特定任務運用的法律問題，依序必須思考的問題如下：

問題一：機器學習「是否」能於特定任務中加以應用？

問題二：機器學習應「如何」在特定任務中加以運用？

問題三：具有「何種品質」的機器學習才能夠在特定任務之中加以運用？

[44] 換言之，如果單純以現行技術下的人工智慧作為討論對象，一般實定法對於人類適用法律所設下的課責措施，就已經足以處理人工智慧的應用問題。也是因為本文設想的討論對象是能力發揮到極致的人工智慧，某些實定法律的界線就沒有辦法有效的處理人工智慧所帶來的問題，而必須如本文一般從理論的層次出發加以討論。

doi:10.6342/NTU202103631

這三個問題彼此牽連、互相影響，且皆具有重要性。抽象的來說，一個任務在社會中的本質以及重要性，決定了機器學習是否能夠運用於特定任務之中，以及機器學習的應用方式；而基於前述論述所生的應用方式，又以具有特定品質的機器學習為前提。問題三問題的釐清，必須將規範內涵形塑為能與科技現實接軌的政策，不僅需要法律管制規範的建構，也必須視現實中的技術內涵發展相應調整。

本論文在論述上僅著眼於問題一以及問題二，並未處理問題三。筆者認為這樣的問題界定是正當的。作為一本臺灣法學院研究生所撰寫的論文，本文設定的閱讀與討論對象尤其是臺灣的讀者以及臺灣的政府。表面上，我國公部門在人工智慧的發展上雖然仍在起步階段，但在台面下正如火如荼的快速發展，並逐漸開始透過人工智慧輔助相關任務的執行。[45]觀察我國發展中的人工智慧類型，多屬於圖像辨識的類型，尚未開始發展到文字探勘、從過往文字的決定中找尋關聯並加以應用的類型。因此，本文所討論的「透過機器學習從事法律適用的任務」，在目前的時空背景下，仍然是未來式。因此，相較於問題三，本文透過釐清問題一與問題二，在此時能夠初步的為人工智慧科技形塑未來發展的方向。

[45] 參見：曹悅華，〈AI影像辨識再加無人機空拍　台北、高雄15路口安全分析一目了然！〉，《聯合報》，2021年9月3日，ogy.de/1181（最後瀏覽日：2021年9月22日）；衛生福利部，〈健保資料人工智慧應用研討會今盛大舉行，展現AI科技應用成果〉，2020年10月27日，ogy.de/dugm（最後瀏覽日：2021年9月22日）；〈影像辨識自動抓！台南設置AI利器 能揪出未戴口罩者〉，《自由時報》，2021年6月23日，ogy.de/s4rt（最後瀏覽日：2021年9月22日）；季大仁，〈防疫創新結合AI人工智慧　熱影像儀偵溫、疫調面面俱到〉，《臺灣好新聞》，2021年6月22日，ogy.de/ct88（最後瀏覽日：2021年9月22日）。

doi:10.6342/NTU202103631

1.5 本文架構

為說明前述的論點以及實際的管制措施，本文的架構安排如下。首先，本文將於**第二章**中，從實際的管制措施出發，說明什麼是「人為介入」，並且釐清這種管制措施背後的事實判斷。之所以從具體的「人為介入」管制措施出發，理由在於這種管制措施是近期在學界中受到廣泛討論、在比較法（歐盟一般資料保護規則）上被付諸實踐的實際的管制措施。但是，即便這種管制措施受到強烈的關注，文獻上對於這種的管制措施背後的事實判斷究竟為何、是否具有正當性，缺乏討論。因此，本文將以歐盟法上對於「人為介入」的制度實踐，分析這種管制措施的內涵、在現實中應如何操作以及釐清這個規範背後的事實判斷究竟為何。

蘊含於「人為介入」管制措施背後最重要的規範意涵是「人類與人工智慧有所不同」，而這種管制措施的正當性，尤其取決於這個規範意涵是否正確。本文將於**第三章**中，聚焦分析人工智慧進行適用法律、作成決定的任務時，與人類是否具有差異，以釐清前述的管制措施是否具有正當性。由於現行的法律系統是以「人類」適用法律為前提加以建立，因此傳統基礎法學中對於「法律適用的內涵」的討論，即是立基於人類適用法律應具有的特色進行論述。本文將以基礎法學中對於「法律適用的內涵」作為基準，與人工智慧適用法律的過程進行比較，以說明人工智慧進行適用法律作成決定將對於現行的法律系統造成的影響。由於本文的目的不在終局性的回答「法律的內涵究竟為何」的難題，僅在於釐清人工智慧適用法律與人類適用法律的差異，因此本文將以法學界對「應如何適用法律」的討論中所形成的相關重要共識作為基礎，據以分析人工智慧在運作上對於法學所造成的影響，以滿足本文論證上的需要。透過本章的分析，本文將指出人工智慧與人類在進行法律適用上，存在著兩個差異：一、無法針對新的案件類型從事法律適用；二、無法因應個案之間的差異，從事法律的續造。

在描述性的分析人工智慧與人類在從事適用法律任務的差異後，本文將於**第四章**中進一步以憲法作為基礎，規範性的分析這兩個差異在憲法下應如何評價，並且進一步說明政府在何種條件之下方能夠合乎憲法要求的使用人工智慧，以從事適用法律、做成決定的任務。在建立了法律續造在憲法法治

doi:10.6342/NTU202103631

原則中的重要性後，本文以此作為基準，一方面將檢視人工智慧是否有不用符合前述要求的例外原因，另一方面分析「人為介入」在實踐上，如何能夠使公部門在使用自動化科技時同時符合憲法的要求，以具體的建立公部門使用自動化科技的正當性條件。為有效的落實「人為介入」的目的，本文主張應揭露「設計階段的重要資訊」。這種以設計階段的各種人為判斷作為透明對象的揭露手段，能夠有效協助公部門中的法律適用者，評估是否推翻或維持人工智慧的建議，符合人為介入的目的。

必須指出的是，由於本文是以人工智慧於公部門中的管制基礎理論為討論的對象，研究的目的主要在於提出供現實規範建構的基礎。因此，具體的規範應該如何設計則非本文討論的重點。本文僅於必要的範圍內，才對於細部的規範設計加以說明。

doi:10.6342/NTU202103631

doi:10.6342/NTU202103631

第二章

「人爲介入」的內涵

本章的研究目的在於分析本文主張所採取的「人為介入」管制措施的內涵，以及蘊藏於這種管制措施背後的規範意涵究竟為何。

在第一章中，本文指出人工智慧的發展目的在於取代人類；在法律的領域中，尤其透過自動化決定取代人類適用法律所作成的決定。弔詭的是，不少學者的倡議，以及諸多國際針對自動化決定的管制架構中，反而採取「人為介入」的管制措施，作為避免自動化決定發生傷害的手段。究竟「人為介入」指的是什麼？能夠達成何種自動化決定管制的目的？

本章將於 2.1 中說明人為介入的內涵。本文將說明人為介入的概念、在實踐上的不同樣態，並且進一步說明人為介入在實際上如何與人工智慧進行互動。

在釐清人為決定與人為介入的內涵後，本章將於 2.2 以歐盟一般資料保護規則為例，說明歐盟資料保護法如何落實這種管制措施的要求。之所以選擇歐盟一般資料保護規則作為討論標的，原因在於歐盟一般資料保護規則作為目前國際間針對自動化決定直接產生管制效力的法規範，規範內容較為細緻，並且討論豐富，有助於釐清人為介入在實際上的操作樣態。本文將指出，歐盟一般資料保護規則透過自動化決定效果產生的拒絕（第 22 條第 1 項）、事後的人為介入（第 22 條第 3 項）及透明權（第 13 條第 2 項第 f 款、第 14 條第 2 項第 g 款以及第 15 條第 1 項第 h 款）等法制措施，有效的落實人為介入的要求。

在釐清歐盟一般資料保護規則對於人為介入的規範內涵後，本文將於 2.3 中嘗試說明蘊含於歐盟一般資料保護規則背後的規範意涵。本文將指出，歐盟立法者針對自動化決定以人為介入加以管制的事實判斷為：一、自動化決定不同於一般的個人資料處理程序，而具有特別管制的重要性；二、自動化科技的運用下，澈底的保留人為決定的可能性；三、自動化決定無法完全取代人為決定。這三種事實判斷，相當於對現行人類決定方式的保留、現狀的

doi:10.6342/NTU202103631

維護，並為人工智慧的使用劃下界限。然而，這條界限的劃定是否正確，仍必須進一步的釐清。

doi:10.6342/NTU202103631

2.1　人爲介入作爲自動化決定的管制手段

自動化決定的目的在於取代人類作成從事適用法律作成決定的任務。然而，為了避免自動化決定所造成的問題，一種管制取徑主張調整這種自動化決定與人類決定者的關係。詳細的來說，這種主張認為應將人類不需要干涉自動化決定結果的關係（human out of the loop）加以改變。除了完全不得使用自動化科技外，一種比較折衷的立場是應將「人類」置於自動化科技的運作過程之中，要求「人類」對於最終產生效果的決定具有影響力，而非任由自動化決定直接產生效力[1]。廣義的解讀此種主張，除了包含人類能夠影響最終單一決定的情形外，自動化決定系統在建置之中，各種關於系統內涵的設計決定也能包括在此種主張之中[2]。但本文此處所討論的僅處理狹義的人類能夠影響最終決定、由人類保留最終作成的狹義的「人為介入（human intervention）」的要求。[3]

從決策者使用人工智慧的方式來說，「人為介入」的管制措施分別為兩種不同的政策選擇。第一種政策選擇，是將人工智慧置於決策過程中輔助者的地位，人工智慧提供建議給人類，由人類參考人工智慧的建議內容作成最終的決定（human in the loop）。在這種類型之中，最終作成決定的主體仍舊是人類，並非人工智慧。[4]本文稱此種類行為「輔助人為決定型」。另外一種政策選擇，則是以人工智慧作為決定的主體，但保留人類能夠介入運作過程

[1] *See*, *e.g.*, Meg Leta Jones, *The Right to a Human in the Loop: Political Constructions of Computer Automation and Personhood*, 47 SOCIAL STUDIES OF SCIENCE 216, 217 (2017); Emily Berman, *A Government of Laws and not of Machines*, 87 B.C. L. REV. 1277, 1283 (2018); Frank A. Pasquale, *A Rule of Persons, Not Machines: The Limits of Legal Automation*, 87 GEO. WASH. L. REV. 1, 46-49 (2019); John Nay & Katherine J. Standburg, *Generalizability: Machine Learning and humans-in-the-loop*, *in* RESEARCH HANDBOOK ON BIG DATA LAW 295-300 (Roland Vogel ed. 2021).

[2] *See* Kiel Brennan-Marquez, Daniel Susser & Karen Levy, *Strange Loops: Apparent versus Actual Human Involvement in Automated Decision-Making*, 34 BERKELEY TECH. L.J. 745, 749-750 (2019).

[3] 除了「人為介入」外，另一種的描述方式則為「受人為決定權」。這兩種描述方式的差異在於前者所包括的範圍較後者廣，除了事先的人為決定之外，也包括了事後由人類推翻自動化決定的可能；後者僅狹義的指涉由人類做成決定的情形。*See* Aziz Z. Huq, *A Right to a Human Decision*, 106 VA. L. REV. 611, 620-629 (2020).

[4] *See* Jake Goldenfein, *Algorithmic Transparency and Decision-Making Accountability: Thought for Buying Machine Learning Algorithms*, *in* CLOSER TO THE MACHINE: TECHNICAL, SOCIAL, AND LEGAL ASPECTS OF AI 48 (Sven Bluemmel ed. 2019); FRANK PASQUALE, NEW LAWS OF ROBOTICS: DEFENDING HUMAN EXPERTISE IN THE AGE OF AI 3-7 (1 ed. 2020).

doi:10.6342/NTU202103631

改變自動化決定效果的可能性。本文稱此種模式為「人為介入保留型」。人為介入保留型的模式中，原則上容許自動化決定的使用，但人類能夠透過不同的方式影響最終的個案決定，例如審查個案中自動化決定的運作是否正當、或保留人類能夠推翻自動化決定的權限（human on the loop）。[5]在實踐上，人為介入會透過程序保障的制度加以落實，例如賦予受決定者能夠透過一定的機制，挑戰人工智慧所作成的自動化決定。[6]比較這兩種類型，人為決定的管制模式對於自動化決定的使用限制較多，將自動化科技侷限於輔助的角色；而人為介入的管制模式，則讓自動化決定居於優先的決定角色，人類負擔事後的擔保責任。

在國際對於人工智慧、自動化決定加以管制的規範之中，不少國際規範採取這種管制措施。例如：經濟合作暨發展組織（Organization for Cooperation and Development, OECD）於 2019 年提出《人工智慧原則》中即以「人類在必要的時候加以介入自動化決定」作為人工智慧符合法治與人權保障的手段[7]；歐盟於 2019 至 2021 年間分別提出的《值得信賴的人工智慧準則[8]》、《人工智慧白皮書[9]》以及統合各會員國關於人工智慧的規範草案[10]，

---

[5] *See* Goldenfein, *supra* note 4, at 48.

[6] *See*, *e.g.*, Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1305-1308 (2007); Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process For Automated Predictions*, 89 WASH. L. REV. 1, 30-32 (2014); Kate Crawford & Jason Schultz, *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms*, 55 B.C. L. REV. 93, 121-128 (2014).

[7] OECD的人工智慧原則共包涵五個子原則：一、透過提升包容性、永許發展以及福祉使人民以及社會受益；二、AI應以符合法治、人權、民主與多元性的方式加以設計並且應該包涵適當的保護措施（例如：**使人在必要的時候加以介入**，以確保公正以及正直的社會）；三、應透過透明以及負責的揭露以確保個人理解基於AI所生的結果，且能對其加以挑戰；四、AI應作用於豐富、安全的方式，並且持續評估以及管控其風險；五、開發、使用以及操作AI的個人或組織應以符合前開原則方式的用AI。OECD Principles on AI, ogy.de/gvhs (last visited: Jul. 1, 2021).

[8] 相較於2018年的歐盟人工智慧策略著重AI的開發，2019年的準則明顯著重焦點於倫理以及法治的要求，其提出八點的要求：**人為監督**、技術健全與安全、隱私與資料治理、透明、多元性不歧視與公正、社會與環境福祉、課責。*See* Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions - Building Trust in Human Centric Artificial Intelligence, COM(2019)168, 4-6 (Apr. 8, 2019), available at: ogy.de/26vz.

[9] 白皮書的內容主要在於提供同時促進AI的運用以及避免AI所造成危害的各種政策選項，以達成以下兩個目的：一、促進AI使用環境的良善，使公部門、私部門、大企業、中小企業以及個人都能夠使用人工智慧；二、創造AI使用的可信任的環境，確保人工智慧的風險不會產生，並且符合歐盟法的價值。其中對於人工智慧的使用設計上，**白皮書中特別要求必須有人為監督的要求**，並提供不同的介入選項。European Commission, White Paper on Artificial Intelligence – A European Approach to Excellence and Trust, COM(2020)65, 2-3, 21 (Feb. 19, 2020), available at: ogy.de/kgps.

[10] 此規則的草案以區分風險的高低作不同的管制方式。其將人工智慧的應用領域區分為不可接受的風險、高

doi:10.6342/NTU202103631

其中皆強調了「人為介入」、「人為監督」的重要性；美國國會政府課責總署於 2021 年提出人工智慧於聯邦政府各部門使用的課責框架中，也說明了人類監督（human supervision）在人工智慧運用上的要求[11]；加拿大政府於 2019 針對公部門內部對於自動化決定的使用所公佈之《自動化決定指令》，區分自動化決定的影響程度，落實不同的人為介入的要求，並且要求說明自動化決定作成的理由[12]；新加坡的個人資料保護總署於 2019 年提出、2020 年加以修訂的模範人工智慧治理框架，同時也要求區分自動化決定的影響程度以對應不同程度的人為介入[13]。以上列舉的人工智慧法制框架，雖然並沒有明確的說明人為介入在人工智慧治理上的目的究竟為何，但可以得知，人為介入是人工智慧管制框架上，不可或缺的一環。

不過，換個角度來思考這種管制措施的發展趨勢，會發現非常奇特的概念衝突。人工智慧之所以被開發以及運用，目的在於取代人類決定，但在規範上卻要求人類必須是決定的主體，或由人類擔保自動化決定所產生的問題。在此規範前提下，自動化決定似乎沒有辦法完全的取代人為決定。但不乏見解指出自動化決定相較於人為決定快速、正確、不會受到偏見的影響[14]。從此

風險、有限的風險、小風險，並且有相應的管制措施，**其中涉及高風險的人工智慧運用，尤其要求必須落實人為監督的要求**，就此可參見草案之第14條的規定。Proposal for a Regulation of the European Parliament and of the Council laying down Harmonized Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending certain Union Legislative acts, COM(2021)206, 3, 51 (Apr. 21, 2021), available at: ogy.de/j92v.

[11] 此課責框架主要針對治理、資料、運用以及監督四個觀點以促進人工智慧使用的課責要求。其中在運用的面向之中，此課責框架強調了必須採取人類監督，Governmental Accountability Office, Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities, 54, available at: ogy.de/zqze (Jun., 2021).

[12] 加拿大的指令僅具有內部的效果。本指令的內容包括自動化決定使用的一般性要求（例如：自動化決定系統的公開近用、設計與監督、安全、資料品質維護等義務）、區分使用領域風險高低的特殊要求（決定過程的說明要求、人類介入、通知要求等特殊性義務），以及公部門若不符合相關要求的法律效果。其中說明決定形成原因的規範請參見6.2.3，要求人為介入請參見6.3.9-6.3.10。Directive on Automated Decision-Making, 2019 (authorized under section 4.4.2.4 of the Policy on Service and Digital under the section 7 of the Financial Administration Act) (Can.), available at: ogy.de/87at (effected on: Apr. 1, 2019).

[13] 此指引的原則有二，原則一是人工智慧作成的決定必須具有可解釋性、透明性以及公正；原則二則應該是以人類為中心。此二原則所推導出的子原則有以下四個：使用人工智慧組織的內部控制要求、**人工智慧輔助決定中人為介入的程度**、資料運用的管理要求、資料使用的利害關係人之間的互動與溝通。Personal Data Protection Commission in Singapore, Model Artificial Intelligence Governance Framework (Second Edition), 30-34, available at: ogy.de/7bra (Jan. 21, 2020)

[14] *See* Huq, *supra* note 3, at 629-650; Cass R. Sunstein, *Governing by Algorithm? No Noise and (Potentially) Less Bias*, DUKE L. J. (manuscript at 2) (forthcoming) (September 15, 2021). Available at: ogy.de/d8t8.

doi:10.6342/NTU202103631

規範的趨勢，不禁讓人好奇「人為介入」究竟是為了達成何種人工智慧上的管制目的？

為釐清此問題，本文將進一步觀察人為介入在實際上應該如何實現。相較於前述尚不具有對外直接拘束力的指引與規範，本文以下進一步分析目前採取的人為介入理念的實際規範—歐盟一般資料保護規則，以釐清蘊含於這種管制措施背後的規範意涵。

doi:10.6342/NTU202103631

2.2　歐盟一般資料保護規則對於自動化決定的管制內涵

本節在簡單的介紹歐盟一般資料保護規則的規範結構後（2.2.1），分別從體系性（2.2.2）與個別性的觀點（2.2.3），說明這個法律架構如何管制自動化決定。

2.2.1　歐盟一般資料保護規則的規範結構與適用範圍

為因應新興科技對於個人資料保護權所產生的影響，歐盟於 2016 年制定一般資料保護規則（General Data Protection Regulation, GDPR; Datenschutz-Grundverordnung, DSGVO）[15]，取代於 1995 年開始適用的資料保護指令（Data Protection Directive; Datenschutzrichtlinie）[16]，並於 2018 年正式施行。新法對於資料保護的程度提升，在形式上直接展現於以下三個變化[17]：一、規範的形式由直接具有拘束會員國而不需要透過內國法程序「規則（Regulation; Verordnung）[18]」，取代過去必須經內國轉換程序的「指令（Directive; Richtlinie）[19]」，調和各成員國的個人資料保護要求；二、罰鍰法定數額提升至資料管理者年營業額的特定比例或者數千萬歐元，且在裁罰上以數額較高者為實際處罰數額[20]；三、適用的範圍除設立於歐盟境內的資料管控者外，尚

[15] Regulation (EU) 2016/679, of the European Parliament and the Council of 27April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1, 88 [hereinafter GDPR].

[16] Directive 95/46/EC, of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, 1995 O.J. (L 281) 31, 50 [hereinafter DPD].

[17] *See* Bryce Goodman & Seth Flaxman, *European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation"*, 38: 3 AI MAGAZINE 50, 51-52 (2017).

[18] 規則，在效力上具有完全的、直接的拘束力。所謂的「直接」指的是規則無庸經過轉換的程序，直接在會員國發生效力。所謂的範圍，在「人」的範圍上，其對於歐盟機構、會員國以及歐盟人民具有直接拘束力；在「物」的範圍上，規則的內容必須完全適用，不得選擇部分適用。參見洪德欽（2015），〈歐盟法的淵源〉，收於：洪德欽、陳淳文編，《歐盟法之基礎原則與實務發展（上）》，頁10，臺北：國立臺灣大學出版中心。

[19] 指令，雖然對於會員國具有拘束力，但對於指令內容實踐的形式以及方式，由會員國自行決定後，轉換成內國法後，方生效力。參見洪德欽，前揭註18，頁10-11。

[20] 歐盟一般資料保護規則對於違反本法的不同情勢分為兩種裁罰，第一種是以最高處以最高歐元一千萬罰鍰或前一會計年度全球營業額百分之二之金額（以較高者為裁罰數額），第二種是以最高歐元二千萬罰鍰或前一會計年度全球營業額百分之四之金額（以較高者為裁罰數額），GDPR, art. 83.

doi:10.6342/NTU202103631

及於任何非設立於歐盟境內但處理歐盟成員國國民個資的資料管理者，而具有域外效力[21]。

關於歐盟一般資料保護規則在事物的適用範圍上，立法者在規則的第 2 條加以限縮。除了排除歐盟法適用範圍與公約以外的情況外，第 2 項第 c 款及第 d 款分別排除了日常個人生活資料利用的情形，以及基於刑事追訴目的所生的資料處理利用行為。不過關於刑事追訴目的所生的資料處理行為並非不受資料保護的規範，歐盟立法者則是透過另外一個 2016/680 指令落實相關刑事過程中的資料保護要求[22]。

對於歐盟一般資料保護規則條文爭議的釐清，除了仰賴學者以及司法實務之見解外，有另外兩種類型的權威解釋依據可以作為參考。一個是附隨於規則前言（preamble）的「序言（recital）」，另一則是依據歐洲資料保護委員會（European Data Protection Board）[23]對於法條爭議闡釋所作成的意見[24]。附隨於歐盟法的序言，目的在於說明規則的制定源由，在規範效力上，其並不具有直接的拘束力，然其作為立法者的意圖，對於歐盟法中文義不清晰的規範，在操作上具有解釋引導的功能。[25]歐洲資料保護委員會由各會員國之資料保護監督機構之代表加以組成，其必須檢視規則適用所生的問題，並且提

[21] GDPR, art. 5.

[22] Directive (EU) 2016/680, of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA, 2016 O.J. (L 119) 89, 131 [hereinafter Directive 2016/680].

[23] GDPR, art. 68.

[24] 隨著歐盟一般資料保護規則的通過，歐洲資料保護委員會於2018年5月25日正式取代依據資料保護指令第29條成立的工作組織（the Working Party of Article 29），成為歐盟一般資料規則下的資料保護監督與管理機關。本文所關注的自動化程序以及透明的要求，過去諮詢工作組織已經發佈兩份的意見，*see* Article 29 Data Working Party, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, 17/EN WP251rev.01 (Feb. 6, 2018) [hereinafter Guidelines on ADM and Profiling]; Article 29 Data Working Party, Guidelines on Transparency under Regulation 2016/679, 17/EN WP260rev.01 (Apr. 11, 2018). 就諮詢小組所作成有關於自動化決定的意見相關闡釋*see* Michael Veale & Lilian Edwards, *Clarity, Surprises, and Further Questions in the Article 29 Working Party Draft Guidance on Automated Decision-Making and Profiling*, 34: 2 COMPUT. LAW SECUR. REV. 398 (2018).關於諮詢工作組織所發佈意見的重要性，*see* Bryan Casey, Ashkon Farhangi & Roland Vogl, *Rethinking Explainable Machines: The GDPR's 'Right to Explanation' Debate and the Rise of Algorithmic Audits in Enterprise*, 34 BERKELEY TECH. L. J. 145, 171-173 (2019).

[25] *See* Tadas Klimas & Jflrate Vaitiukait, *The Law of Recitals in European Community Legislation*, 15: 1 ILSA J. INT'L. & COMP. L. 61, 73, 76 (2008).

doi:10.6342/NTU202103631

供適用上的意見。[26]過去，這個組織所作成的意見，經常受到歐盟法院參考，並作為裁判的依據，指引司法實務的見解形成。以下說明歐盟一般資料保護規則之中對於自動化決定的「特別」規範內容。

2.2.2　歐盟一般資料保護規則對於自動化決定的管制內涵

歐盟一般資料保護規則以「個人資料保護」為規範目的，因此就本文所欲討論的「自動化決定」而言，只要決定過程之中使用了個人資料，即構成此規則中的「資料處理」，而應適用本法的一般規定。然而，相較於一般的資料處理行為，歐盟一般資料保護規則的立法者透過其它規定，特別的對於「自動化決定」進行管制。

詳言之，歐盟立法者對於自動化決定的管制，特別想要達成以下兩個目的：一、避免資料主體的權利受到自動化決定侵害；二、透過體系性的框架建立，糾正自動化決定可能發生的問題。[27]在權利保障的面向之中，歐盟立法者為了要避免資料主體受到自動化決定的特殊侵害，因此透過資訊告知權（義務）的建構讓資料主體知悉自動化決定的存在，資料主體能夠藉由相應的資訊獲取權理解所面臨的自動化決定的內涵[28]。此外，資料主體能夠選擇不受到自動化決定的拘束，請求以人為的方式作成決定。但立法者並非一概的禁止自動化決定的使用。在滿足例外條件的情形中，仍舊容許自動化決定的使用，但附加一定的事後程序保障的要求，賦予資料主體能夠請求人為介入、表達意見甚至對於自動化決定進行爭訟等措施，作為底線要求，達成資料主體權利的保護目的。[29]換言之，歐盟立法者透過透明手段、受人為決定以及人為介入的方式保障受自動化決定下決定者的權利。

---

[26] GDPR, art. 70.

[27] *See* Margot E. Kaminski, *Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability*, 92 S. CAL. L. REV. 1529, 1537-1552, 1586, 1595 (2018).

[28] GDPR, art. 13-15.

[29] GDPR, art. 22.

doi:10.6342/NTU202103631

此外，在框架性結構的面向之中，立法者則透過行為準則[30]、認證制度[31]或資料保護官[32]的措施，確保各公部門組織與私部門產業中的資料處理內部規範，符合歐盟一般資料保護規則的要求。特殊的是，在具體應採取的措施選擇上，則是由資料管控者對於其本身所使用的資料處理行為從事資料影響評估（data protection impact assessment, DPIA）[33]判斷自身所從事資料處理行為所造成的風險大小，並自行決定所因應的保護措施[34]。

在權利面向的相關規範中，歐盟立法者分別以透過資料主體對於直接蒐集資料的資料管控者對於自動化決定的受告知權（第 13 條第 2 項第 f 款）、資料主體對於間接蒐集資料的資料管控者對於自動化決定的受告知權（第 14 條第 2 項第 g 款）、資料主體向資料管控者請求提供自動化決定使用相關資訊的近用取得權（第 15 條第 1 項第 h 款）以及對於自動化決定的拘束拒絕權（第 22 條）等四個規範規範自動化決定。以下分別說明此各該權利的規範上依據以及其內涵。

2.2.3 管制自動化決定規範的要件釐清

2.2.3.1 自動化決定拘束拒絕權與人爲介入權

歐盟立法者在歐盟一般資料保護規則之中透過第 22 條的「自動化個人決定，包括特徵剖繪」作為管制中心。[35]原則上賦予個人得拒絕受到自動化決定的法律效果所拘束的權利，例外容許使用自動化決動的情形之中，則賦予請求人為介入的機會。[36]本法的目的在於，避免自動化決定對於受決定人民自由、

[30] GDPR, art. 40-41

[31] GDPR, art. 42.

[32] GDPR, art. 37.

[33] GDPR, art. 35.

[34] Margot E Kaminski & Gianclaudio Malgieri, *Algorithmic Impact Assessments Under the GDPR: Producing Multi-layered Explanations*, 11 INT'L DATA PRIV. LAW 125, 132-133 (2021). 學者稱此種管制模式為「雙元的管制（Binary Governance）」，透過公私協力的方式建立保護資料主體的管制框架，Kaminski, *supra* note 27, at 1582.

[35] GDPR, art. 22.

[36] *See* Isak Mendoza & Lee A. Bygrave, *The Right Not to be Subject to Automated Decisions Based on Profiling*, *in* EU INTERNET LAW: REGULATION AND ENFORCEMENT 85 (Tatiana-Eleni Synodinou, et al. eds., 2017).

doi:10.6342/NTU202103631

權利可能發生的特殊風險[37]，以及確保人類能夠控制並為決定加以負責[38]。其具體條文如下[39]：

第 22 條 自動化個人決定，包括特徵剖繪

I. 資料主體應有權不受單獨基於自動化處理（包括特徵剖繪）所做成而對其產生法律效果或類似重大影響之決定所拘束。

II. 若該決定有以下情形，第 1 項不予適用：

(a) 為締結或履行資料主體與控管者間之契約所必要者；

(b) 由受規範之歐盟法或會員國法明文授權，且定有適當之保護措施以保障資料主體之權利及自由及正當利益者；或

(c) 基於資料主體的明確同意。

III. 於第 2 項第 a 款與第 c 款之情形，資料管控者應實施適當之保護措施，使資料主體至少有權得就資料管控者部分為人為的介入、表達其意見以及爭執該決定以保障其權利及自由及正當利益。

---

[37] *See* PAUL VOIGT & AXEL VON DEM BUSSCHE, THE EU GENERAL DATA PROTECTION REGULATION (GDPR): A PRACTICAL GUIDE 180 (1 ed. 2017)；中文文獻可參見張陳弘、莊植寧（2019），《新時代之個人資料保護法制：歐盟GDPR與臺灣個人資料保護法的比較說明》，頁208，臺北：新學林。

[38] *See* Lee A. Bygrave, *Minding the Machine v2.0: The EU General Data Protection Regulation and Automated Decision Making*, *in* ALGORITHMIC REGULATION 249-250 (Karen Yeung & Martin Lodge eds., 2019).

[39] 一般資料保護規則第22條的英文原文如下：

Article 22 Automated individual decision-making, including profiling

1. The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.

2. Paragraph 1 shall not apply if the decision:

(a) is necessary for entering into, or performance of, a contract between the data subject and a data controller;

(b) is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or

(c) is based on the data subject's explicit consent.

3. In the cases referred to in points (a) and (c) of paragraph 2, the data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision.

4. Decisions referred to in paragraph 2 shall not be based on special categories of personal data referred to in Article 9(1), unless point (a) or (g) of Article 9(2) applies and suitable measures to safeguard the data subject's rights and freedoms and legitimate interests are in place.

doi:10.6342/NTU202103631

IV. 基於第 2 項所為之決定不得以第 9 條第 1 項之特種個人資料為基礎，但構成第 9 條第 2 項第 a 款或同項第 c 款且存在保障資料主體權利以及自由以及正當利益的適當之保護措施時，不在此限。

這個規定並非首次見於 2016 年的一般資料保護規則之中。早於 1995 年資料保護指令第 15 條中即有類似規定[40]。資料保護指令第 15 條是參考 1978 年法國資料保護法加以制定[41]。這個規範的特色在於其並非基於個人資料保護下「公正合理資訊處理原則[42]（Fair Information Practice Principles, FIPPs）」的規範內涵；這個規範主要是針對資料的「應用」情形加以規制。[43]換言之，相較於其它資料保護法的規定，本條並不以個人資料「處理過程」作為規範的對象，而是以基於資料應用所產生具有形式外觀的單一「決定」作為規範對象。[44]

延續指令第 15 條所制定的規則第 22 條，在規範上與舊法仍有若干差異。其中最大的差距在於適用範圍，第 22 條第 1 項所規範的「自動化程序」範圍明顯**大於**資料保護指令。[45]指令所規範的對象，僅限於目的在預測的「特徵剖繪」，以此種資料處理為唯一的自動化決定規範類型；而規則規範的對象將

[40] DPD, art. 15.

[41] *See* Lee A. Bygrave, *Article 22*, *in* THE EU GENERAL DATA PROTECTION REGULATION (GDPR): A COMMENTART 529 (Christopher Kuner, et al. eds., 2020).

[42] 此原則是建立於告知同意基礎上的資料保護管理機制，其內涵至少包括了以下五個原則：告知原則（Notice/Awareness principle）、選擇及同意原則（Choice/Consent principle）、接近及參與原則（Access/Participation principle）、完整及安全原則（Integrity/Security principle）、實踐與救濟原則（Enforcement/Redress principle）。就此內涵中文文獻可參見，林子儀（2015），〈隱私權法制的新議題：監控與隱私自我管理〉，收於：國立臺灣大學法律學院編，《第五屆馬漢寶講座論文彙編》，頁96-98，臺北：財團法人馬氏思上文教基金會。

[43] *See* Diana Sancho, *Automated Decision-Making under Article 22 GDPR:Towards a More Substantial Regime for Solely Automated Decision-Making*, *in* ALGORITHMS AND LAW 140 (Martin Ebers & Susana Navas eds., 2020).

[44] 類似德國、我國行政程序法中的行政處分。Vgl. von Lewinski, in: BeckOK Datenschutzrecht, 32. Aufl., 2020, DS-GVO Art. 22, Rn. 27; Kumkar & Roth-Isigkeit, Erklärungspflichten bei automatisierten Datenverarbeitungen nach der DSGVO, JZ 2020, 277, 278.

[45] *See* Bygrave, *supra* note 38, at 252; Maja Brkan, *Do Algorithms Rule the World? Algorithmic Decision-making and Data Protection in the Framework of the GDPR and Beyond*, 27 INT. J. LAW INFO. TECH. 91, 96-97 (2019).

doi:10.6342/NTU202103631

特徵剖繪[46]作為一種例示的類型，涵納其他可能產生法律效果或重大影響的「自動化決定」[47]。

然而，此項規範在要件以及效果上仍有若干不明確之處。以下分別就本條第 1 項要管制的標的（A.）、第 1 項的管制手段（B.）、不同項之間的關係（C.）加以說明。

A.　第22條第1項的管制對象

首先，第 22 條第 1 項要求個案的「決定」是單獨基於自動化處理所作成。在規範對象上，其以形式上的最終「決定」為規範標的，因此若僅是單純自動化決定的「程序」，若最終並未做成決定，並不在管制的範圍。

此外，受到管制的標的是「單獨」基於自動化處理所作成的決定。所謂的「單獨（based solely on）」指的是**無人為的介入**。「單獨」作為要件，其立法目的在於避免無人為介入的程序中可能產生的特殊風險，亦即預設了「人的決定」以及「自動化決定」有所不同。「單獨」作為 1995 年資料保護指令即存在的要件，其範圍究竟為何，在舊法上的解釋即有所爭議。若狹義解釋「單獨」，此時任何參與決定過程中的「無意義」人為決定，即會排除此規範的規制。因此使用自動化決定者，可以在自動化程序中，納入「橡皮圖章」式的人為介入，而豁免於此規定的規制，並且在結果上大幅的降低此條規範的適用可能性[48]。然而，目前多數的見解，皆認為應採取目的性的解釋，主張所謂的「單獨」應視介入的人是否「有能力影響決定」的結果而定。若有人為介入，但該介入者並不具有能力影響該決定結果時，這種情形仍舊該當本項中的「單獨」；相反的，如果介入者有能力影響自動化決定的結果時，則不該當此處的單獨，理由在於此時相當於人為決定，並不會發生立法者預設自動化決定會產生的特殊風險。歐洲資料保護委員會也是採取此種見解。[49]

---

[46] 在一般資料保護規則之中，立法者於第4條第4項中明確定義特徵剖繪（profiling）的要件。特徵描繪在一般資料保護規則中所包含的範圍極度廣泛，其要件有三：一、以個人資料為處理對象；二、以自動化程序為之；三、目的用以評斷自然人。GDPR, art. 4 (4).

[47] *See also*, Bygrave, *supra* note 41, at 530.

[48] *See* Gianclaudio Malgieri & Giovanni Comandé, *Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation*, 7 INT. DATA PRIV. LAW 243, 251 (2017).

[49] *See* Article 29 Data Working Party, Guidelines on ADM and Profiling, at 20-21. 除此之外英國資訊專員辦公室也

doi:10.6342/NTU202103631

除了決定是獨立基於自動化程序所為之外，該決定必須對於資料主體發生「法律效果」或者「與法律效果相當的重大影響」。所謂的「法律效果」指的是對於資料主體產生法律資格的變更，其所指涉的內涵包括公法以及私法上各種權利的干預[50]。在文義上，雖然本條包含公、私部門利用自動化程序所做成的一切決定，不論效果是正面或者負面，但在解釋上本條僅在處理發生「負面」影響的決定[51]。而「與法律效果相當的重大影響」相對於「法律效果」，其所指涉的行為是不造成法律資格的改變，但是在個案中，對當事人產生強烈、長期等差別對待的事實效果時，即屬之。[52]

從以上的分析，可以得知第 22 條所欲處理的標的是無人為介入，且產生法律效果或者是事實上重大影響的「自動化決定」。參酌本章 2.1 中的論述，依據人工智慧在決策中的角色，可以區分為作為輔助者的情形，或居於決定者的地位。而第 22 條處理的對象是單獨基於自動化處理所生法律效果的個別「自動化決定」，可以得知第 22 條專門處理居於決定者地位、無人為介入的全自動化決定。換言之，「作為輔助者的人工智慧」並非第 22 條所處理的對象；也因此，輔助型人工智慧的使用，並不會使第 22 條發生法律效果。

B. 第22條第1項的法律效果

針對自動化決定，本條在法律效果上，容許受決定者「無庸受該決定所拘束」。所謂的無庸受到該自動化決定所拘束，指的是受決定者能夠拒絕受到自動化決定的效果所及。設想一個情形，公部門以自動化決定的方式對於

採取此種見解。*See* Information Commissioner's Officer, Feedback request – profiling and automated decision-making 19 (Apr. 28, 2017) ("…we think it is intended to cover those automated decision-making processes where a ***human exercises no real influence on the outcome of the decision***… (italic added)").

[50] 學者普遍認為自動化決定對於公權利的限制（干預）會構成此處的「法律效果」。然而發生爭議者，主要在於私法領域中，利用自動化決定禁止特定人締結契約是否構成此處的法律效果？常見的例子是雇主透過自動化決定系統排除投遞履歷應徵者。否定見解認為基於私法自治，且民事權利主體不因為拒絕締約而受到權利侵害；肯定見解認為特定脈絡中若拒絕締約造成權利侵害則屬之。Vgl. Lewinski, (Fn.44), Rn. 31. 本文基本上認為基於私法自治原則，單純拒絕締約並不生權利的侵害，但在特定脈絡中，要約者具有締約請求權時（如該等契約之承諾者有強制的締約義務，如基礎性服務），此時透過自動化決定而否准該契約的締結的確會產生法律效果，而構成本項的要件。

[51] Vgl. Lewinski, (Fn. 44), Rn. 33.

[52] 此標準由歐盟資料保護指令第29條諮詢工作組織提出，*see* Article 29 Data Working Party, Guidelines on ADM and Profiling, at 21. 相類似見解例如Gianclaudio Malgieri與Giovanni Comandé認為該影響若造成不平等的影響、利用消費者的弱點加以操縱時即該當，*see* Malgieri & Comandé, *supra* note 48, at 253.

doi:10.6342/NTU202103631

特定當事人作成決定，而當事人主張此權利。此時造成公部門以自動化決定方式作成的決定不生效力。公部門若仍有義務作成該決定，僅能以「人為決定」的方式使法律效果實際的產生。基於此原因，本條在法律效果上可以稱為請求「受人為決定」[53]。

雖然此規範置於第三章的「資料主體之權利」，且在文字上以「權利」作為描述，然而學者間對於其法律效果，在本質上究竟為資料主體的權利？或是使用自動化程序決定的資料管控者應承擔的一般禁止性規定？有重大的爭議。[54]詳言之，若將此規範解釋為資料主體的權利，此規範之意旨即相當於「原則上資料管控者得使用自動化決定，但於資料主體行使此權利時，則例外不得為之」；相反的，若將此規範解釋為一般性的禁止義務時，其規範意旨即相當於「原則上禁止使用獨立的自動化程序，毋待資料主體加以行使」。這兩種不同解釋方法，對於資料主體的權利保障、決策者是否能夠使用自動化決定產生重大的影響。

此爭議於資料保護指令時期即已存在，但由於「指令」的體例性質，不需要在歐盟層次解決，而是留待歐盟成員國在內國法化第 15 條的過程中加以形成[55]，因此成員國有採取權利模式制定者[56]、有採取禁止義務[57]，甚至有區分決定主體為公部門或者私部門，而分別採取禁止義務或權利的混合立法[58]。但是此爭議在規則規範時期，即有加以釐清的必要；因為「規則」在歐盟法中的目的即在於整合歐盟各國的法制，不容許各成員國有不同的資料保護體制。

採取「權利說」者，大多從本法的文義以及規則制定的沿革加以說明。首先，系爭規範以不受自動化決定的「權利」稱之，且其規範於此規則第三章的「資料主體的權利」。[59]再者，若本規範欲以禁止論的形式加以規範，立

---

[53] *See* Sancho, *supra* note 43, at 147; Huq, *supra* note 3, 622-624; Brkan, *supra* note 45, at 17.

[54] *See*, *e.g.*, Sandra Wachter, Brent Mittelstadt & Luciano Floridi, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, 7 INT. DATA PRIV. LAW 76, 94-96 (2017); Brkan, *supra* note 45, at 8-9.

[55] *See* Mendoza & Bygrave, *supra* note 36, at 86.

[56] 如瑞典於1998年所制定的資料保護法的第29條即採取權利模式，Bygrave, *supra* note 41, at 529.

[57] 如比利時於1992年所制定的資料保護法的第12條及採取禁止義務模式，*id*. at 529.

[58] 如義大利於2003年的資料保護法，區分決定者的對象是公部門還是私部門而加以應對。如果是司法決定（判決）或者是行政行為，則採取禁止義務模式立法；若是私部門為之，則採取權利模式立法。

[59] Mendoza & Bygrave, *supra* note 36, at 85.

doi:10.6342/NTU202103631

法者應該採取本條第 4 項「應禁止（shall not; dürfen nicht）」的用語[60]。就此以觀，立法者應是有意採取「權利」模式的制定方式。此外，部分學者從立法制訂的源由觀之，認為所繼受的法國法原先本是一種禁止規範，而在進入指令的時代後，立法者在立法的過程之中有意的改變文字採取「權利說」的主張。[61]

而採取「禁止說」者，則從目的解釋以及條文之間的規範關係加以說明。首先，相較於本法同章的其它權利，本條的「權利」內容不在於積極的請求（to）特定作為，而是不受特定作為（not to be subject），本質上是一種消極的權利。而本規則之中的其他權利也不乏消極權（例如第 13 條、第 14 條的資料通知權），因此解釋為禁止義務並不當然違反文義。再者，考量同條第 2 項的豁免規定，邏輯上當事人不可能同時同意以自動化決定的方式作成決定（第 c 款），又行使第 1 項的權利不受自動化決定的拘束。且從強化資料主體權利的觀點出發，將此規定解釋為一「原則上禁止」的義務規範，方具有殺傷力，能澈底排除自動化決定的風險，而不待資料主體加以主張。[62]此外，若將第 22 條第 1 項的法律效果解釋為必須主動行使的權利，將架空同法同條第 2 項第 b 款、第 3 項例外允許自動化決定但必須符合的程序要求（即「適當的保護措施」）。[63]詳言之，在「權利說」下，若當事人並未行使第 1 項的權利時，決策者就能毫不保留的使用自動化決定，並不會受到立法者所課與的程序要求義務。[64]

特殊的是，補充一般資料保護規則適用刑事例外的 2016/680 指令中[65]，其中第 11 條對於自動化決定於刑事追訴中的情形，明確的採取了「禁止說」的立場，除非會員國立法授權容許並且該授權必須同時提供人為的介入的保障措施，否則原則禁止自動化決定的使用。[66]而歐洲資料保護委員會於指引中明

[60] *Id*. at 86.

[61] Luca Tosoni, *The Right to Object to Automated Individual Decisions: Resolving the Ambiguity of Article 22(1) of the General Data Protection Regulation*, 11 INT. DATA PRIV. LAW 145, 148-152 (2021).

[62] Mendoza & Bygrave, *supra* note 36, at 86.

[63] *Id*. at 87.

[64] *See* Wachter, Mittelstadt & Floridi, *supra* note 54, at 95; Brkan, *supra* note 45, at 8.

[65] 本指令適用於對於刑事追訴的過程之中，而並未適用於司法審判的情形。

[66] Article 11 of Directive 2016/680: Automated individual decision-making

1. Member States shall provide for a decision based solely on automated processing, including profiling, which

doi:10.6342/NTU202103631

確的指出，基於加強資料主體對於個人資料的控制強化，本條雖然名為「權利」，但在解釋上其相當於**一般性的禁止規定**，不論資料主體是否加以行使，而序言中第 71 點的用語也可顯示立法者制定為一般禁止規範的意圖[67]。歐盟會員國荷蘭海牙地方法院（Hague District Court）於 2020 年 5 月 2 日作成判決中也依循前述工作組織的見解，認為此規範應為禁止規範[68]。

C.　第22條第2項與第3項的內涵

針對第 1 項的「不受自動化決定拘束」的效果，同條第 2 項設有三種情形的豁免條件，若符合三種情形中之一種情形時，決定者仍舊可以使用自動化決定，但必須提供程序的保障。

第一個豁免條件「締結或履行契約所必要[69]」係指自動化決定為契約之成立以及履行所不可或缺。所謂的「必要」在解釋上必須考量是否有其它能達成契約締結或履行的目的對於資料保護侵害較小之手段。然而，若狹義解釋「必要」一語，將無自動化決定得透過本款加以豁免，蓋在現實中並無契約之締結或者履行必須透過自動化決定方得為之。第二款的「會員國或者歐盟法授權允許[70]」豁免要件則是彰顯了一般資料保護規則僅為一框架性的規範，歐盟或者會員國得透過立法加以例外容許自動化規定。不過規則仍對於例外容許的情形設定的框架性的規範，要求必須建立適當的保護措施。[71]而第三個

produces an adverse legal effect concerning the data subject or significantly affects him or her, ***to be prohibited*** unless authorised by Union or Member State law to which the controller is subject and which provides appropriate safeguards for the rights and freedoms of the data subject, at least the right to obtain human intervention on the part of the controller (*italic added*).

2. Decisions referred to in paragraph 1 of this Article shall not be based on special categories of personal data referred to in Article 10, unless suitable measures to safeguard the data subject's rights and freedoms and legitimate interests are in place.

3. Profiling that results in discrimination against natural persons on the basis of special categories of personal data referred to in Article 10 shall be prohibited, in accordance with Union law.

[67] Article 29 Data Working Party, Guidelines on ADM and Profiling, at 19-20.

[68] Rb. Den Haag 2 mei 2020, C/09/550982/HA ZA 18/388, ¶ 6.35-6.36 (ECLI:NL:RBDHA:2020:1878).

[69] GDPR, art. 22(2)(a).

[70] GDPR, art. 22(2)(b).

[71] 依據學者的研究，目前歐盟會員各國依據此款而特別立法容許自動化決定總共有八國，分別為比利時、荷蘭、法國、德國、匈牙利、斯洛維尼亞、澳洲、愛爾蘭，英國於脫歐前亦有就此特定制定規範Gianclaudio Malgieri, *Automated decision-making in the EU Member States: The Right to Explanation and other "Suitable Safeguards" in the National Legislations*, 35: 5 COMPUT. LAW SECUR. REV. (2019).

doi:10.6342/NTU202103631

「基於資料主體明確同意[72]」豁免要件，在資料主體明確表達同意受自動化決定的情形下，容許使用自動化決定。「明確同意」為資料保護規則立法所特別強調的重點。規則第 4 條第 11 款條要求「同意」必須是資料主體出於自由意志，對於所同意對象受明確告知，並且積極、確實的為特定事項的同意。因此，對於自動化決定使用的同意，資料主體亦必須對自動化決定的存在、內涵完全知悉，並且為積極同意的意思表示。

除了此三個豁免要件之外，對於使用特別種類個人資料為自動化決定時，若符合「當事人同意為之且無歐盟法或會員國立法之明文禁止」或「當事人雖然於法律上或事實上無法為同意但對於保護當事人的權利具有必要性」其中之一個條件，且具有適當的保護措施時，即得例外允許基於特別各種資料所為的自動化程序。

在符合「契約例外」以及「一般同意例外」的情形，規則設下了如同第二項 b 款的程序保障要求，其要求自動化程序的使用者必須建立足以保障資料主體權利的「適當的保護措施」，相較於第二項 b 款（授權例外）以及第四項（特種個資例外），規則尚對於程序的內涵進行特定，必須使資料主體至少能夠「就資料管控者部分為人為的介入、表達其意見以及爭執該決定」。就此內涵，規則的序言中的前言（recital）第 71 點對此有以下的說明：「…在任何情況中，該處理必須有適當的保護措施，包括應給予資料主體的特定資訊以及獲得人為介入的權利、表達其（資料主體）意見的權利、對於經過評估後做成決定**獲得說明的權利**以及爭執決定的權利。…」

前言在兩方面澄清了「適當的保護措施」的內涵。第一方面是應適用程序要求的情形。前言指出「在任何情形中」，皆應該包含例示的相關措施，亦即即便在第 22 條第 2 項第 b 款以及同條第 4 項立法者並未要求「適當保護措施」內涵的情形中，也應該適用第 22 條第 3 項所明文的程序要求。第二方面是程序要求的內涵，本點除了重複說明第 22 條第 3 項的內涵外，還創設了自動化決定後的個案說明權。然而，前言亦僅止於此，並未繼續的闡明何謂說明權、說明權的內涵為何。

[72] GDPR, art. 22(2)(c).

doi:10.6342/NTU202103631

總結以上的說明，第 22 條所針對的對象是「自動化決定」，僅處理利用自動化科技作成決定的情形，並不處理輔助型的自動化科技。在個案中，如果受決定者在面對產生重大影響的自動化決定，且沒有第 22 條第 2 項的例外時，第 22 條第 1 項的效果就會因此發生，決定者不得使用自動化決定，只能使用人為決定，此時法律並不禁止透過自動化科技的輔助作成決定。但若在個案之中，具有第 22 條第 2 項的例外時，決策者能夠使用自動化決定，受決策者不能夠行使第 1 項的拒絕權，但決策者因此也具有第 22 條第 2 項第 b 款與同條第 3 項的義務，提供「人為介入」等程序保護措施。[73]

2.2.3.2　涉及自動化決定重要資訊的告知權/義務與近用權

針對自動化決定，除了有前開第 22 條的拘束拒絕權加以規範，規則尚分別以第 13 條第 2 項第 f 款、第 14 條第 2 項第 g 款、第 15 條第 1 項第 h 款三個條文，創設了關於「告知權」與「近用權」的規範。

這三條規範目的在於強化「透明（transparency）」、「公正（fairness）」與「合法（lawfulness）」所賦予資料主體的權利，使其知悉其個人資料的行使狀態、並加以控制[74]。基於同一個目的，這三個規範差異在於前二者是課予資料管控者告知義務，分別要求資料管控者於直接取得個人資料（第 13 條第 2 項第 f 款）時，或間接取得個人資料時（第 14 條第 2 項第 g 款）告知關於自動化決定的相關資訊，第 15 條第 1 項第 h 款則是賦予受決定的對象能夠主動的行使接近使用權（the right to access）向資料管控者請求一定的資訊。

在制定沿革上，第 13 條第 1 項第 f 款以及第 14 條第 2 項第 g 款是歐盟一般資料保護規則中全新的規範，而第 15 條第 1 項第 h 款的資料近用取得權的

---

[73] 一個法條操作上的問題是，如果個案之中使用自動化決定，並且符合第22條第2項的例外容許使用自動化決定的豁免情形，但決策者並未提供相應的程序保護措施時，此時第22條第1項的「拒絕效果」是否會產生，而導致禁止決策者使用自動化決定？針對此問題，法律上並沒有規範的很清楚。就法條的文義而言，第22條第2項本文指出符合同項下三款的任一情形，即不得行使第22條第1項的拒絕權，而立法者並未將第22條第3項的程序保護義務內容以類似第b款的立法方式置於第2項之中，因此即便沒有提供相關的人為介入的措施，決策者仍舊能夠使用自動化決定。因此地如果是第b款的「會員國或者歐盟法授權允許」的情形，由於第b款的內容。另外一種可能的解釋方式，則是以第2項必須與第3項聯結加以解釋，自動化決定例外的容許使用必須以提供相關程序保護措施為前提。本文認為應採取後面的這種解釋方式，理由在於本文認為**人類決定與自動化決定的不同**，足以正當化人為介入作為使用自動化決定的必要條件。關於這個論點，本文將於後續加以論證。

[74] VOIGT & BUSSCHE, *supra* note 37, at 88.

doi:10.6342/NTU202103631

前身即為資料保護指令的第 12 條第 a 款第 3 目，賦予資料主體得請求關於特徵剖繪「自動化處理的邏輯」的相關資訊[75]。由於第 13 條第 2 項第 f 款與第 14 條第 2 項第 g 款的內容相同，僅翻譯第 13 條第 2 項第 f 款[76]與第 15 條第 1 項第 h 款[77]的條文內容如下：

第 13 條 直接從資料主體處取得個人資料時應提供之資訊

II. 除第 1 項規範之資訊外，管控者應於取得個人資料之時提供資料主體以下更進一步的必要資訊，以確保公正且透明的處理：

(f) 第 22 條第 1 項與第 4 項所指稱之自動化決定（包括特徵剖繪）的存在，至少於前開情形中，包括運作邏輯、以及該處理程序對於資料主體的重要性以及可預見結果的有意義資訊。

第 15 條 資料主體的資訊接近使用權

I. 資料主體應有權向管控者獲得確認是否有其個人資料正在被處理，若其資料正被處理，其有權接近使用個人資料以及以下資訊：

---

[75] Gabriela Zanfir-Fortuna, *Article 15*, *in* THE EU GENERAL DATA PROTECTION REGULATION (GDPR): A COMMENTARY 463 (Christopher Kuner, et al. eds., 2020); DPD, art. 12(a).

[76] 一般資料保護規則第13條第2項第f款的英文原文如下：

Article 13 Information to be provided where personal data are collected from the data subject

…

2. In addition to the information referred to in paragraph 1, the controller shall, at the time when personal data are obtained, provide the data subject with the following further information necessary to ensure fair and transparent processing:

(f) the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.

[77] 一般資料保護規則第15條的英文原文如下：

Article 15 Right of access by the data subject

1. The data subject shall have the right to obtain from the controller confirmation as to whether or not personal data concerning him or her are being processed, and, where that is the case, access to the personal data and the following information:…

(h) the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.

doi:10.6342/NTU202103631

(h) 第 22 條第 1 項與第 4 項所指稱之自動化決定（包括特徵剖繪）的存在，至少於前開情形中，包括運作邏輯、以及該處理程序對於資料主體的重要性以及可預見結果的有意義資訊。

就個人資料保護的觀點而言，第 13 條、第 14 條本文的受告知權（告知義務）是資料保護的第一層的保障，第 15 條本文的資訊接近使用權是資料保護的第二層保障。資料主體本得透過此二種權利獲得以下三種資訊：一、獲得特定資訊是否被處理的答覆；二、取得資料處理過程的細節；三、確切被處理的資料[78]。此處針對自動化決定的三個特別規範，則是對於「涉及自動化決定的資料」同時課予告知義務，並且賦予主動近用的權利。

在資訊提供的門檻上，本法於文字中強調「至少於」符合第 22 條的自動化決定時，才有提供相關資訊的義務。由於歐盟一般資料保護規則對於違反本法的規定有高額的裁罰，因此從此條文可以得知僅於涉及第 22 條中的「自動化決定」時，資料管控者才有提供資訊的法律義務[79]。

這三個規範要求資料管控者應提供以下三種資訊給受決定的對象：一、自動化決定存在的事實；二、涉及決定系統邏輯的有意義資訊；三、說明自動化決定對於資料主體可能產生影響的有意義資訊。

在第一個資訊上，自動化決定存在的事實為一「是/否」的答案，並無適用上的疑義。因此，資料管控者必須針對本身是否使用第 22 條的自動化決定告知受決定的對象。若此答案為是，資料管控者隨之具有後面兩種資訊的提供義務。

在第二個「涉及系統邏輯的有意義資訊」中，資料管控者應該提供何種資訊，學者指出此**並不**指涉所使用的特定演算法、自動化決定的程式碼，所謂的「有意義」應採取有利於「資料主體」的解釋，提供該運作程序的建置基準[80]。而歐洲資料保護委員會也採取類似的見解，認為所謂的涉及決定系統

[78] Zanfir-Fortuna, *supra* note 80, at 453 (Article 15).

[79] Vgl. Paal/Hennemann, in: DS-GVO GDSG, 2. Aufl., 2018, DS-GVO Art. 13, Rn. 32.

[80] *See*, *e.g.* Zanfir-Fortuna, *supra* note 80, 429-430; Andrew D. Selbst & Julia Powles, *Meaningful Information and the Right to Explanation*, 7 INT. DATA PRIV. LAW 233, 236 (2017).

doi:10.6342/NTU202103631

邏輯的有意義資訊，其內涵即為決定背後的基準；此外資料管控者並不用揭露或者說明整個演算法，僅要提供足以使資料主體理解之所以獲得決定的理由即可。[81]

在第三個「對於資料主體可能產生的影響」的資訊中，歐洲資料保護委員會要求必須明確的以可理解的方式，以具體的例子說明自動化決定「可能」發生的效果，例如使用自動化決定判斷保費的保險業者，應清楚的告知受決定者，何種行為會導致高額的保費、何種行為會降低保費。[82]

一個比較不清楚的問題是，這三個規範與第 22 條權利的行使之間是否具有一定的關聯性？在法條之間的關聯上，從這三個關於資訊請求權條文中「至少於第 22 條第 1 項中」的文義可以得知，資料管控者應告知的資訊，與資料主體得請求之資訊，應該是涉及第 22 條規範的特定「自動化決策」內涵的資訊。而透明權下應提供的資訊內容上包括「自動化決策的存在」、「自動化決策對於資料主體的重要性以及預設結果的有意義資訊」以及「涉及運作邏輯的有意義資訊」三種資訊，這三個資訊是涉及資料主體是否行使第 22 條第 1 項與第 3 項權利的重要資訊。[83]就實際而言，資料主體若欲主張「受人為決定」或者後續的「人為介入」，至少必須知道自動化決策的存在，以及該自動化決策是否將造成重大的影響。[84]就此，可以得知第 13 條第 2 項第 f 款、

[81] Article 29 Data Working Party, Guidelines on ADM and Profiling, at 25.

[82] Article 29 Data Working Party, Guidelines on ADM and Profiling, at 26.

[83] 若觀察歐盟一般資料保護規則的制定緣由，可以發現此種關聯性有其道理。詳言之，歐洲執委會於2012年1月25日提出歐盟一般資料保護規則的第一版立法草案中，草案第20條「基於特徵剖繪的措施（Measures based on Profiling）」的第4項指出資料管控者於系爭草案第14條的資訊告知義務，其內涵包括「自動化決策存在的事實」，以及該程序對於主體「可能產生的效果」。就此法條的安排可以得知，關於自動化決策的資訊告知義務，其目的即在於輔助行使同條第1項的人為介入權。雖然關於資訊「近用取得權」的規範，歐洲執委會仍舊留待於普通規定（草案第15條）加以處理，並且最終呈現於規則的第15條第1項第h款，但由於第15條的文字與第13條與第14條，並且條文中強調了本條與第22條的關聯，第15條的功能理應按照第13條以及第14條的方式進行說明，即所謂的「涉及運作邏輯有意義的資訊」即為相關輔助資料主體決定是否應行使第22條的使用權。Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), COM (2012) 11 final (Jan. 25, 2012).

[84] 論者或許會指出，第13條第2項第f款、第14條第2項第g款以及第15條第1項第h款的「涉及運作邏輯的有意義資訊」並不如同另外兩個要件與第22條第1項相互對應，因此強加建立此三規範與第22條的連結似乎過度牽強。但這個主張有問題之處在於其根本上的誤解了「涉及運作邏輯有意義資訊」的功能與意義。「涉及運作邏輯有意義資訊」的功能在於使資料主體知悉自動化決策的決策標準為何，其是否與人為決策有所落差。詳言之，第22條第1項在效果上為「人為介入」的權利，而資料主體之所以希望人為介入即是考量並且希望到人為決策的結果與自動化決策的結果並不相同，而此種評估則有賴於知悉「涉及運作邏輯的有

doi:10.6342/NTU202103631

第 14 條第 2 項第 g 款以及第 15 條第 1 項第 h 款存在的目的，在於滿足第 22 條第 1 項所行使的前提所存在，透過「揭露」的透明方式，使資料主體知悉「自動化決策的存在」、「自動化決策對於資料主體的重要性以及預設結果的有意義」，並決定「第 22 條第 1 項」的權利是否應加以行使。

2.2.3.3　定位不明的說明權

除了前面兩種權利之外，學者之間具有重大爭議的是，歐盟一般資料保護規則之中是否賦予資料主體「說明權」，能夠請求決定者說明特定自動化決定作成的理由。在 2.2.3.1 的論述中，可以得知說明權的爭議起源於此法第 22 條第 3 項「適當保護措施」與前言第 71 點中的內涵。這個權利爭議的焦點分別在於權利的內涵為何、規範依據何在以及其在自動化決定的管制框架之中具有何種功能。本文認為，說明權在自動化決定中是否能具有功能，必須釐清此權利的規範依據以及目的為何，方能分析這權利對於自動化決定的管制是否具有功能。因此以下主要說明本權利的內涵為何，以及整理學者間對於此權利的依據所生的爭議。

「說明」指的是說明者（explainer）就向被說明者（explainee）解答「為何問題（why-question）」的過程。[85]說明的目的在於建立特定已發生事件與特定原因之間的合理關聯。[86]在方式上，說明者對於待說明事項（explanandum）採取逆推推論的方式，從眾多的理由之中決定一說明，並將其提供與被說明者。[87]被說明者於接受到說明後，說明人與被說明人即生一定的社會互動，而該互動視所提供的「說明」為何，會達成不同的目的。[88]說明的內容視情境以及其欲達成的目的而有所不同，例如「日常的說明」著重於特定決定的說明；而「科學的說明」著重於一般關係形成的說明。[89]從此處對

意義資訊」。一言以蔽之，第13條至第15條的權利目的在於幫助資料主體行使第22條第1項以及第4項的權利。

[85] Tim Miller, *Explanation in Artificial Intelligence: Insights from the Social Sciences*, 267 ARTIFICIAL INTELLIGENCE 1 (2019).

[86] Vgl. Käde & Maltzan, Die Erklärbarkeit von Künstlicher Intelligenz (KI): Entmystifizierung der Black Box und Chancen für das Recht, CR 2020, 66, 67 f.

[87] Miller, *supra* note 85, at 7.

[88] *Id*. at 34.

[89] *Id*. at 6.

doi:10.6342/NTU202103631

於「說明」內涵的一般性闡述可以得知說明至少有以下兩個特徵：一、在時點上，說明必然是一個「事後」的溝通過程，對於已發生事件所提供的資訊；二、說明的內涵依據目的而有所不同，可能說明的是特定事件的發生原因，亦有可能是對於整體事件的概括理由。從以上的說明可以釐清說明權的內涵在眾多學者心中之想像。

學界目前對於歐盟一般資料保護規則中說明權的討論，主要依據說明的「時點」以及說明的「對象」區分出以下三種類型的說明[90]：自動化決定作成前的系統運作邏輯的說明（ex ante, system functionality explanation）、自動化決定作成後的系統運作邏輯的說明（ex post, system functionality explanation）以及自動化決定作成後的個別決定的說明（ex post, specific decision explanation）。[91]此外，從第 13 條第 2 項第 f 款、第 14 條第 2 項第 g 款以及第 15 條第 1 項第 h 款的文字，可以得知請求的資訊包括「涉及運作邏輯的有意義資訊」；且尤其是第 15 條的權利不論在決定之前或者決定之後皆能由資料主體發動，關於自動化決定作成前的系統運作邏輯的說明，及自動化決定作成後的系統運作邏輯的說明，在現行法上至少都有明確的法律基礎，唯獨「自動化決定作成後的個別決定的說明」在規範之上並無清楚的依據，因此「說明權的爭議」可以說是涉及「**自動化決定作成後的個別決定說明請求權在歐盟一般資料保護規則的根據為何**」的爭議。

關於此爭議，可以區分為三種見解。第一種見解認為第 22 條第 3 項的「適當保護措施」可以作為說明權的請求權基礎；第二種見解則是以第 13 條第 2 項第 f 款、第 14 條第 2 項第 g 款的「涉及運作邏輯的有意義資訊」作為請求權基礎；第三種見解認為能夠透過第 15 條第 1 項第 h 款的「涉及運作邏輯的有意義資訊」作為請求權基礎。

就第一種見解來說，序言中的說明權，指涉的對象是「決定之後」獲得的「獨立決定」的說明，資料管控者有義務就其決定個案提供決定作成的理

[90] 之所以僅具有三種類型而不包括「自動化決定前的個別決定的解釋」的類型，是因為此種類型並不存在。

[91] 有學者認為技術現實上自動化決定所使用的模型本身即蘊含了個案決定作成的原因，並且不會因為決定之前或者決定之後的時間點有所不同，因此認為此三分類並不具有重要性，*see* Selbst & Powles, *supra* note 80, 239-241. 但本文認為此分類仍然具有重要性，因為對於被說明人而言，其無法理解完整模型說明中的哪一個部分構成了個案說明，且對於被說明人而言「整體說明」以及「個案說明」可能具有不同的意義。

doi:10.6342/NTU202103631

由。就此內涵而言，其屬於自動化決定的事後個別決定的說明[92]。但是第 22 條第 3 項所推導出的說明權是否具有拘束決定者的效力，而使決定者必須負擔說明義務存有疑義。主要理由在於從文字上來理解，可以得知說明權僅為「適當措施」的一種類型，資料主體自然可以選擇「其它」能夠達成人為介入、表達意見以及對決定進行爭執的其它措施，因此即便資料管控者不賦予「說明權」，但提供其它符合前開程序要求的機制，即不違反本條的規定。若欲使「適當措施」具有課予「說明義務」的規範效力，一種可能的解釋方式是釐清立法者對於「適當措施」的內涵為何。若立法者的主觀意圖即在以「適當措施」指稱說明權，則該條所衍生的說明權自具有法拘束力。但從歷史解釋的觀點[93]來分析，可以得知這個論點不具有說服力。不論在規則修法的過程之中[94]，可以發現有明確的將「說明權」置於草案的條文之內，但後續刪除並置於序言中的紀錄。據此可之，立法者於本法制訂時並未打算創設一具有「拘束力」的說明權。[95]

第二種見解認為規則的第 13 條第 2 項第 f 款與第 14 條第 2 項第 g 款中「涉及運作邏輯的有意義資訊」能夠作為「自動化決定作成後的個別決定的說明」的請求權基礎。然而，這個請求權基礎的問題在於文義上與「事後的」要求有所不符。第 13 條與第 14 條在性質上為資料管控者對於自動化決定存在

---

[92] Kumkar & Roth-Isigkeit, (Fn.44), S. 281.

[93] 相類似的一般資料保護規則制訂史探究， *see, e.g.,* Brkan, *supra* note 45, at 6-7; Wachter, Mittelstadt & Floridi, *supra* note 54, at 81-82.

[94] 歐洲議會全會於2014年3月14號針對歐盟一般資料保護規則的草案進行一讀的表決，其對於原草案中第20條的規定，涉及內容變更的修正部分如下：一、其將原先對於特徵剖繪的拘束拒絕「權」明確改為一「禁止規範」；二；強化對於「特種資料」所為「特徵剖繪」的禁止要求；三、其課予資料管控者一義務，要求在涉及法律效果以及事實上影響重大的特徵剖繪，應有「人為的評估」，且人為的評估的內容包含「**人為評估後提供的說明**」，此即為本條中的「適當的措施」。Position of the European Parliament adopted at first reading on 12 March 2014 with a view to the adoption of Regulation (EU) No .../2014 of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), EP-PE_TC1-COD (2012) 0011 (Mar. 12,2014), 148-150.

此次歐洲議會的一讀是「說明權」正式出現於歐盟一般資料保護規則的制定歷史之中。然而，皆下來歐盟理事會對於本法的修正，首先將本條第1項「禁止規範」改為「權利」，並且其刪除了歐洲議會所提出的說明權版本，並將說明權的文字從法規移至前言序言的第58點之中。Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)- Preparation of a general approach, 9565/15 (Jun. 11, 2015), 32, 106.

[95] *Contra* Malgieri & Comandé, *supra* note 48, at 255. Malgieri與Comandé認為前言第71點作為補充第22條文義不清楚之處可以作為說明權行使的依據。

doi:10.6342/NTU202103631

的通知義務，其通知時點在於資料蒐集的「當下」，而蒐集資料的當下必然在決定時點之前，因此相較於「說明權」所具有的事後性質，此二條僅得為「事前」的告知；此外，自「可預見（envisaged）」的文義觀之，其係以「未來」該程序可能存在的影響作為提供告知內容，更可以說明這兩條的告知義務，無法作為說明義務/權利的來源。[96]

至於第三種見解認為第 15 條第 1 項第 h 款的資訊近用權能夠作為「自動化決定作成後的個別決定說明」的請求權基礎。而第 15 條第 1 項第 h 款在法條文字上雖然與第 13 條、第 14 條相同，但由於權利本質有所不同，其於行使上即有時點上的差距。詳言之，不同於第 13 條與第 14 條受通知權（通知義務）的性質，第 15 條在性質上為「資訊接近取得權」，資料主體得於決定作成之前或者決定作成之後行使，因此在性質上並不具有行使的時間限制。[97]因此，此權利如果於決定後行使即該當說明權的「事後」要求。但本條在文義上以「可能（envisaged）結果」一取向於未來的用語形塑權利的內涵，限縮本條僅能於決定之「前」加以行使而無法該當所謂的「事後的說明權」，且在本條的舊法中（個人資料保護指令第 15 條）在行使的內涵上僅限於運作邏輯而未包括獨立決定。[98]此外，歐洲資料保護委員會明確的指出第 15 條中的「可能的結果」**並非**特定決定的個別說明，僅是涉及自動化決定的一般性資訊，諸如：自動化決定中所考量的因素、每個因素的比重，等其它有助於資料主體挑戰決定的資訊。[99]

由於目前歐盟法院尚未就此條規範有任何的判決，前開爭議尚未能妥適的化解。本文認為前面眾多學者之間的見解，大多過於注重單一法條上文字

[96] *See* Wachter, Mittelstadt & Floridi, *supra* note 54, at 82. 不過Brkan認為「行使時點」的主張僅有在「直接蒐集」的第13條第2項地f款才能成立，在第14條第2項第g款的「間接蒐集」的情形無法成立。理由在於，間接蒐集的告知義務中，不同於第13條第1項的要求必須於取得資料之「當下」為告知 ，第14條資料管控者於取得資料後合理期間內為告知即符合規則的要求，而合理期間並不長於一個月。因此在現實上，可能存在資料管控者間接取得個人資料，並以此為基礎作成自動化決定，而後在依據第14條第2項第g款加以告知資料主體。在此情形中，由於第14條第2項第g款也有「事後」提供的可能性。*See* Brkan, *supra* note 45, at 23.

[97] Malgieri & Comandé, *supra* note 48, at 255.

[98] *See* Wachter, Mittelstadt & Floridi, *supra* note 54, at 83-84, 87. 反對見解如Malgieri與Comandé認為能夠以條文之中「涉及對於資料主體的重要性以及預設結果的有意義資訊」作為請求權基礎取得個案的決定。Malgieri & Comandé, *supra* note 48, at 258.

[99] Article 29 Data Working Party, Guidelines on ADM and Profiling, at 27.

doi:10.6342/NTU202103631

多義性的理解與說明，並且嘗試將「自動化決定作成後的個別決定的說明權」置入法條之中。[100]但這樣的討論方式對於釐清「說明權」的目的而言，是無意義的。因為這些不同規範所欲達成的目的並不相同，不同規範下的「說明權」皆會有不同的樣貌。因此，重要的反而是影響說明權內容的標準究竟應該是什麼。比較恰當的討論方式毋寧應該是確認所涉及法條的目的，藉此進一步推論透過該法條能夠請求何種內容的說明。

若以歐洲資料保護委員會的意見為基礎，我們可以得知「自動化決定作成後個別決定形成原因說明」的請求權基礎並非第 13 條至第 15 條的規定，而第 22 條第 3 項亦未明文要求提供此種請求權。既然「說明權」的爭議起源於前言第 71 點連結本法第 22 條第 3 項的內容，釐清說明權內涵的依據反而應該視立法者於本項中所要求的「人為的介入、表達其意見以及爭執該決定」的目的。因此，如果有說明權的話，說明的內涵必須釐清個別決定所形成的原因，讓受決定者能夠請求人為介入，分析特定個別的自動化決定決定中何處有問題、對於有問題之處表達意見，以及請求救濟。

2.2.4　小結

從以上的說明，可以得知歐盟一般資料保護規則針對自動化決定的管制，採取了多元的管制模式。一言以蔽之，本法分別綜合採取資料保護影響評估、法令遵循的稽核、外部獨立機關的監督、透明、受人為決定以及人為介入避免自動化決定所造成的影響。其中，針對受決定者的權利保障，立法者以涉及自動化系統的資訊告知義務與近用權為輔助手段，促成受人為決定的要求，原則上使個人得直接受人為決定，例外透過程序保障的要求，以人為介入作為擔保的措施，避免受決定者因自動化決定而受到不合理的影響。

[100] *See also* Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORDHAM L. REV. 1085, 1107 (2018).

doi:10.6342/NTU202103631

2.3 歐盟一般資料保護規則立法者的事實判斷

從 2.2 的分析，我們可以得知歐盟一般資料保護規則對於自動化決定的管制內涵。基於此管制內涵，若我們以「人為介入」為中心，我們可以進一步的推論歐盟立法者對於自動化決定的事實判斷。

首先，歐盟資料保護規則區分了一般的資料處理行為以及自動化決定，除了一般關於個人資料保護的規範之外，立法者特別透過第 13 條第 2 項第 f 款、第 14 條第 2 項第 g 款、第 15 條第 1 項第 h 款及第 22 條的規定針對「自動化決定」進行系統性的特別管制。這樣的規範模式，在自動化社會即將來臨的背景下，可能有兩種意涵。第一種意涵是作為迎接自動化決定全面來臨之前的過渡規範，為了避免自動化決定社會與現行人為決定社會之間的銜接困難，透過保留人為決定、人為介入的可能，漸進式邁向自動化決定的社會。但本文認為，基於歐盟資料保護一般規則作為「資料保護」的框架性意涵，目的應不在於社會轉型的過渡銜接，而在於積極的保障資料主體的權利；此外，從前述法律規範沿革的介紹可以得知，現行法第 22 條並非全新的規範，而是往昔資料保護指令時期，即存在的規範。因此，合理的解釋本規範，目的應在保障資料主體免於自動化決定所造成的不利影響。這些特別針對「自動化決定」規制的不同規範中，還隱含了自動化決定相較於一般的個人資料處理行為會對於資料主體產生不一樣的影響，因此必須透過特別的規範對於自動化決定以不同的方式加以管制。因此，我們可以得出歐盟一般資料保護規則的立法者，對於自動化決定的第一個事實判斷：

(1)自動化決定具有特別管制的必要性。

明確化管制的對象之後，立法者在面對自動化決定所產生的問題時，選擇以第 22 條的「人為介入」作為管制手段以及第 13 條第 2 項第 f 款、第 14 條第 2 項第 g 款及第 15 條第 1 項第 h 款「資訊告知權（義務）」與「資訊近用權」作為管制手段，使受自動化決定的對象，能夠透過資訊的請求，理解自動化決定的內涵，並且進一步的選擇是否行使本法第 22 條第 1 項的受人為決定。此外，立法者為了貫徹這種以「人」為決定主體的規範意涵，於例外容許使用自動化決定的情況中，在第 22 條第 3 項之中創設後續人為介入、表達意見、爭執決定正當性、甚至是說明權等後續保障，透過賦予個人對於自

doi:10.6342/NTU202103631

動化決定挑戰的機會，以達到受決定者權利的保障。雖然事後的程序要求在保障效果上，弱於決定一開始的受人為決定[101]，且使受決定者承擔先受自動化決定拘束的風險，但立法者在此選擇仍舊貫徹了受人為決定的要求。換句話來說，不論是原則上禁止使用自動化決定的情形，或者例外容許使用自動化決定的情形，「人類」皆必須是決定形成的主體。就此，我們可以導出第二個事實判斷：

(2)自動化科技的運用下，應該澈底的保留人為介入的可能性。

更進一步的著眼於「人為介入」的管制手段，姑不論「人類在運作過程中」的功能為何、是否能夠有效的達成目的，人為介入的機制，存在一個重要的前提，就是必須有「人」存在，以落實人為決定的可能。可以進一步的說，全面使用自動化決定的社會，與立法者的想向有所牴觸，因為全面使用自動化決定的社會，將沒有「人」的存在能夠落實人為介入的規範要求。「人為介入」作為一種管制措施，在自動化決定應用的時代之中，表面上是一種全新的要求，但實際上，這種要求主張反而是對於現狀保留的要求，要求原有的決策機制繼續存續；新興決策方式的出現，也不代表舊的決策方式就因此必須加以淘汰。本文認為，立法者之所以要維持現狀的理由，在於立法者認為「人為決定」與「自動化決定」兩種方式，在本質上有所不同。就此，我們可以推論歐盟立法者的第三個事實判斷：

(3) 因為人為決定與自動化決定有本質上的不同，所以反對全面性的使用自動化決定。

然而，本文透過解釋所分析的三個事實判斷是否正當，需要再進一步釐清。

首先，本文認為第一個規範意涵在科技進展的現今是不證自明的。相較於一般性個人資料的蒐集、處理、利用行為，自動化決定透過運用巨量資料之間的關聯性並加以決定發生個案效果的方式，已經不是傳統個人資料保護法所處理的問題。[102]傳統基於資訊隱私之個人資料保護法處理的是個人對於個

[101] *See* Tosoni, *supra* note 61, at 154-156.

[102] *See*, *e.g.*, danah boyd & Kate Crawford, *Critical Questions for Big Data: Provocations for A Cultural, Technological, and Scholarly Phenomenon*, 15 INFORMATION, COMMUNICATION & SOCIETY 662 (2012); Omer Tene & Jules Polonets, *Judged by the Tin Man: Individual Rights in the Age of Big Data*, 11 J. ON TELECOMM. HIGH TECH. L. 351 (2013); VIKTOR MAYER-SCHONBERGER & KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL

doi:10.6342/NTU202103631

人資料的自主控制的問題，但在自動化決定的問題上已非個人資料自主控制的問題，而是第三人以資料內涵的知識對於個人加以分析的結果。因此自動化決定當然有特殊管制的必要性。

然而，第二個以及第三個規範意涵，在現今人工智慧蓬勃發展的情況下，顯的有些過時以及老舊。「人為介入」作為一種自動化決定的管制方式，內涵並不在於賦予受決定者能夠要求決定者作成特定「內容」的決定，而是要求決定者應該以特定的「方式」作成影響他人的決定；此外，這種決定方式的要求，並未要求決策者提供新的決策方式，僅在於要求維持現狀、以現行的方式作成決定，並且以事後的「人為介入」的底線擔保，在最小程度內的落實了這樣的要求，維繫了當代社會之中以「人」作為決定主體的現狀，要求公部門以及私部門的決策方式不得隨著科技的進步而完全改變。這兩個規範意涵，一方面承認了自動化決定與人為決定的差異，另一方面對於此差異進行了規範上的評價，認為人為決定能夠有效的避免自動化決定所產生的問題。但我們仍舊不清楚的是自動化決定與人為決定究竟差距在哪，這種差距具有何種重要性。

人為介入作為一種維持現狀的管制模式，我們應該繼續質問的問題應該是，蘊含於這種管制措施背後的事實判斷是否正當。第二個與第三個規範意涵是否正當的標準，取決於自動化決定與人類決定是否有差異？在何種方面有差異？此種差異具有何種規範上的重要性？唯有釐清自動化決定與人為決定之間的差異，以及這些差異在一定法秩序下的重要性，才能正確的判斷這種管制措施是否具有正當性。

***

總結以上的說明，本章首先介紹了人為介入的規範類型，並且以歐盟一般資料保護規則作為分析的標的，並且釐清了人為介入規範模式背後所蘊含的事實判斷。將目光轉移離開學說上充分討論的「說明權爭議」，本文透過分析歐盟一般資料保護規則的規範內涵，釐清蘊含於背後的立法選擇，並指出歐盟立法者所謂的人為介入指的是當代決定方式的維持，並且立法者隱約

TRANSFORM HOW WE LIVE, WORK, AND THINK (1 ed. 2013).

doi:10.6342/NTU202103631

的認為自動化決定無法完全的取代人類的決定。在新興科技時代下，這種事實判斷是否正確，則是下一章要釐清的問題。

doi:10.6342/NTU202103631

doi:10.6342/NTU202103631

第三章

人類與人工智慧在法律適用上的差異

本文於第二章說明了歐盟一般資料保護規則立法者蘊含於「人為介入」管制模式背後的事實判斷。本章將進一步檢討這些事實判斷是否正確。本章將聚焦於第二波人工智慧，分析人工智慧與人類在從事法律適用時是否真的不同？若有不同，如何不同？其不同是否影響我們的價值選擇？其不同之處是否足以說明必須維持以「人類」作為決定中心的必要？

事實上，人工智慧的使用已經產生了一些不同於人為決定的法律問題[1]。這些爭議之所以發生，一部分可歸因於技術上的障礙尚未完全克服[2]；另一部分的原因是人工智慧在運作原理上就存在的特質所導致。技術性的障礙在科技持續的發展下，有朝一日應可克服；但因為運作原理所生的特性，只要使用此種技術就會如影隨形。因此，若要根本的處理人工智慧與人類決定有何不同，尤其應著眼於基於人工智慧運作原理所產生的問題。

此外，比較人類以及人工智慧是否具有差異、乃至於人工智慧是否能夠取代人類，必須著重一個特定的面向（任務類型），將人類與人工智慧在這個特定面向中的表現結果相互比較。[3]由於本文探討的對象是公部門適用法律作成的個別決定，因此本文將集中於「適用法律」的任務中，分析人工智慧與人類有何差異。

---

[1] Bradford K. Newman, Adam Aft & Yoon Chae, Recent Developments in Artificial Intelligence Cases, BUSINESS LAW TODAY (Jun. 16, 2021), ogy.de/6wm6 (last visited: Jul. 23, 2021).

[2] 本文所謂的因為「技術上的障礙」所造成的問題諸如因為「訓練資料不足」、「訓練資料不正確」、「分類標準有問題」、「資料特徵選擇有問題」等在「設計」自動化決定階段中，因為沒有特定規範準則要求資料以及程式應符合的標準所造成的問題。而這些問題在現今尤其產生了重大的問題，*See* KATE CRAWFORD, THE ATLAS OF AI: POWER, POLITICS, AND THE PLANETARY COSTS OF ARTIFICIAL INTELLIGENCE 89-149 (1 ed. 2021).(chapter 3 & chapter 4)

[3] 針對人工智慧的相關議題討論，最廣泛且引人深思的提問毋寧是「人工智慧是否能夠取代人類」。但這樣的問法，過度寬泛，且沒有辦法聚焦問題的核心。理由在於現行的人工智慧是「窄的人工智慧（narrow Artificial Intelligence）」，並非通用型的人工智慧，僅能解決「特定」的任務。在這樣的技術現實下，比較務實的提問反而應該是「人工智慧是否能取代『特定的』人類任務」。

doi:10.6342/NTU202103631

本章首先將說明人工智慧在技術上如何形成（3.1）。本章將指出人工智慧分別由人類的判斷、資料以及演算法加以構成，而資料以及演算法將形塑人工智慧在表現上的效能。

在釐清人工智慧的技術內涵後，本章將說明公部門中能夠如何運用人工智慧、達成何種人類無法達成的結果（3.2）。本章將指出人工智慧之所以受到公部門的青睞，主要的原因在於人工智慧具有兩種能力：第一種能力是「模仿」，人工智慧能夠以過去的案例為樣本，快速且一致的模仿人類的法律適用行為；第二種能力是「發現資料之間的關聯性」，達到社會控制的目的。基於「資料關聯性」在本質上並非因果關係，使用上與以理性為基礎的法律適用有所不符。本文以下將檢討「模仿型」的人工智慧與人類適用法律有何差異。

為比較人工智慧與人類在適用法律上的差異，本文將勾勒人類從事法律適用任務的過程（3.3）。本章將指出，將抽象法律與具體個案連結的法律適用行為，主要包括兩個任務，「特定法律的個案適用」與「特定法律本身正當性的檢驗」。這兩個任務的目的，在於使法律系統中單一、抽象、預先制定的規範，能夠適切的處理社會中所發生的各種、具體、未來發生的個案。換言之，法律的適用並非僵化的機械性操作，而是視個案的差異而有不同的適用結果，必要時必須突破現行法律的框架，達到典範變遷、法律續造的目的。

在建立起法律適用的內涵後，本章將進一步檢討人工智慧前面兩種能力的定位（3.4）。本文以下分別檢討人工智慧的運作過程無法理解、文義理解具有缺陷以及無法打破常規三個特性是否構成人工智慧與人類之間的差異。本文認為模仿型的人工智慧雖然具有黑箱的性質，但法律論證並不要求提供決定的「原因」，僅要求提供決定的「理由」，而模仿型的人工智慧有能力模仿出人類進行法律適用的理由，因此黑箱的特性並非人類與人工智慧之間的重要差異。本文認為模仿型人工智慧與人類在法律適用上最大的差異，在於無法對於新的案例進行法律論證，以及進行法律的續造。

doi:10.6342/NTU202103631

3.1　人工智慧的技術本質

以資料為基礎的人工智慧，在原理上是以透過演算法針對巨量資料進行訓練，經由提取蘊含於資料之中的規則，並加以運用於現實，達成人類指定的目的。人工智慧的建置步驟如下[4]：

定義目標→蒐集並且處理資料→透過演算法得出模型→調校→應用

從此步驟加以觀之，會發現人工智慧至少包括三個重要的特徵：人類的定義、資料以及演算法。以下從技術面說明自動化決定系統是如何建置，並且說明這三種特色如何影響人工智慧的表現。

人工智慧的出現，目的在於協助，甚至取代人類，解決過去應由人類解決的社會問題。因此，在人工智慧系統建置的第一步驟，就是由人類設計者告訴人工智慧應該從事何種任務。[5]若要讓人工智慧能夠進行人類給予的任務，則必須以人工智慧能夠理解的方式形塑任務的內涵。例如：在自駕車的建置之中，將不要發生人員傷亡的任務化約為避開障礙物。因此，機器學習的系統建置者在設定任務時，必須找到一個能夠代表該任務的「效用函數（utility function）」，讓電腦能夠理解必須從事的任務內涵。[6]在前述自駕車的系統建置之中，代碼找尋的任務是容易的；相比之下，其它社會中必須依賴脈絡加以判斷，或無法量化的任務中，此步驟即產生困難。例如，如何用量化指標，正確的描述刑事被告的再犯風險即有重大的困難。人工智慧的建置者僅能透過獲取領域的知識，盡可能正確的形塑效用函數的內涵。[7]

在完成目標的定義後，系統的建置者接下來的任務就是蒐集並且處理資料。資料在人工智慧系統建置中的地位，主要在於提煉出能夠應用的模型，

---

[4] *See* David Lehr & Paul Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, 51 U.C. DAVIS L. REV. 653, 670-702 (2017); Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 677-694 (2016); Tal Z. Zarsky, *Transparency in Data Mining: From Theory to Practice*, *in* DISCRIMINATION AND PRIVACY IN THE INFORMATION SOCIETY 307-309 (Bart Custers, et al. eds., 2013).

[5] *See* Lehr & Ohm, *supra* note 4, at 672-677; Barocas & Selbst, *supra* note 4, at 677-680.

[6] 不過在機器學習學界是以反面表述，稱為「損失函數（loss function）」，並希望盡可能的縮小損失函數的數值。See STUART RUSSELL & PETER NORVIG, ARTIFICIAL INTELLIGENCE: A MODERN APPROACH 687 (4 ed. 2021); Lehr & Ohm, *supra* note 4, at 675. 刑事審判系統系統中關於再犯率應如何定義，尤其產生爭議，*see* Jessica M. Eaglin, *Constructing Recidivism Risk*, 67 EMORY L. REV. 59, 75-78 (2017).

[7] *See* Lehr & Ohm, *supra* note 4, at 672-677; Barocas & Selbst, *supra* note 4, at 677-680.

doi:10.6342/NTU202103631

因此資料除了要能作為「訓練資料」提煉模型之外，也必須作為「測試資料」測試模型能夠應用的樣本，確保模型在不同情況中的可應用性。[8]建立在歷史會不斷重演的原理，建置者必須確保訓練資料以及測試資料的內涵與過去的歷史能有足夠的吻合。因此建置者必須確保有足夠的資料、資料對於目標具有正確性以及資料的整體具有一般性[9]。相關的措施例如：除去過度特異的資料以保持未來的模型具有一般性、偵測到錯誤或缺漏的資料並且加以補足或者刪除[10]。

在完成資料蒐集以及處理的作業後，建置者必須斟酌所定義的目標，選擇設計適當的演算法針對資料加以訓練。詳言之，不同的演算法能夠達成不同的任務，因此建置者必須透過考量結果的呈現、偏誤比率、演算法本身的可理解性、演算法的可介入性以及資料的內涵選擇足夠恰當的演算法訓練出足以使用的模型[11]。

最後，在模型生成到最後落地使用的過程之間，建置者必須持續的就訓練出的模型進行調校與評估，調整演算法以及資料的各種參數，以製作出最能夠有效達成定義目標的模型，並於最終加以使用。[12]

觀察上述關於人工智慧的建置過程，可以發現特定人工智慧的表現，分別受到以下三種內涵重要的影響：人為的判斷、資料以及演算法。整個人工智慧製作的過程之中，建置者必須作成大大小小的判斷以及定義，包括：定義目標的方式、選擇所蒐集資料、切分訓練資料以及測試資料、確保資料的品質、選擇演算法以及評估模型…等系統建置的因素，建置者本身的知識以及意識形態都會影響決定結果的呈現。[13]此外，以資料為基礎的人工智慧，最知名的特色就是「垃圾進、垃圾出！（Garbage in, garbage out!）」：資料的品質決定了人工智慧表現的品質，也因此人工智慧的建置花費最多的時間在

[8] *See* Lehr & Ohm, *supra* note 4, at 684-688.

[9] *See id*. at 680.

[10] *See id*. at 681-683.

[11] *See id*. at 688-695.

[12] *See id*. at 4, at 695-701.

[13] 相關的人類知識形塑可能導致歧視的結果。*See*, *e.g.*, Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1262 (2007).但通常都是無意識的偏見所造成的結果。*See*, *e.g.*, Anupam Chander, *The Racist Algorithm* 115 MICH. L. REV. 1023 (2016).

doi:10.6342/NTU202103631

處理資料，使資料能夠精準的反應現實，避免人工智慧無法在現實中加以運用[14]。最後，人工智慧之所以能完成有效能的模型的原因在於演算法的使用，透過演算法的運作找到資料之間的關聯性，並且加以應用。

本文透過以上的說明勾勒出人工智慧因為「人為判斷」、「資料」以及「演算法」三種特色所形成的技術本質，但此技術本質的說明似乎沒有辦法說明為何人工智慧被廣泛的使用，並且有取代人類的聲浪。因此本文將繼續說明人工智慧的能力，以及這些能力如何輔助公部門執行任務。

[14] *See* Sandra G. Mayson, *Bias In, Bias Out*, 128 YALE L. J. 2218, 2224 (2019).

doi:10.6342/NTU202103631

3.2 人工智慧的兩種能力

人工智慧的使用目的有兩種，一種是技術性的（technological），用以協助人類從事特定的任務；另外一種目的則是科學性地（scientific），用以回答人類不知道的問題。[15]而這兩種使用的目的，分別造就了人工智慧的兩種不同的核心能力：「模仿」以及「洞見」。[16]前者指的是人工智慧能夠學習並複製人類的決策行為；後者則是能夠從資料之中發現人類無法察覺得「洞見」。本節將說明，這兩種特色如何協助公部門從事法律適用的任務。

3.2.1 模仿型的人工智慧

自從有電腦開始，法律自動化適用一直是學界努力的方向。最早對於法律自動化適用的發展，是以「將法律視為程式」的觀點加以發展，不過此種發展方向並未成功。將法律視為程式的觀點指的是從形式上觀察法律運作的過程，可以發現法律就像是公式，只要將個案（數值）置入法條（公式）之中，最後就會得出答案。即便這樣的思維有一些問題[17]，早期對於法律自動化適用的實踐卻是著眼於此思維。在這種思維之下，其目標即在建立電腦能適用的法律適用模型。將擬適用的法律規範予以程式化，隨後將個案輸入程式之中，而得出該個案適用特定法律規範的法律效果。在 1970 年代的資訊學者即開始對於此主張付諸實踐，相關領域的研究以「人工智慧與法律（Artificial Intelligence & Law）」稱之，[18]其目的在建置出能夠從事能解決法律問題的「法律論理的運算模型（Computational Models of Legal Reasoning, CMLRs）」。[19]此種推論方式可以稱為演繹法（deduction），推論的基礎是

[15] *See* MARGARET A. BODEN, ARTIFICIAL INTELLIGENCE: A VERY SHORT INTRODUCTION 2 (2018).

[16]參見邱文聰（2021），〈亦步亦趨的模仿還是超前部署的控制？AI的兩種能力和它們帶來的挑戰〉，收錄於《智慧新世界？人文社科學人眼中的AI》；邱文聰（2020），〈第二波人工智慧知識學習與生產對法學的挑戰—資訊、科技與社會研究及法學的對話〉，收於：李建良編，《法律思維與制度的智慧轉型》，頁137-168，臺北：元照

[17] *See* L. Karl Branting, *Data-centric and Logic-based Models for Automated Legal Problem Solving*, 25 AI & L. 5, 7-11 (2017).

[18] 對於此領域的介紹，以及「人工智慧與法律」或「法律與人工智慧」的用語，中文文獻參見陳弘儒（2020），〈初探目的解釋在法律人工智慧系統之運用可能〉，收於：李建良編，《法律思維與制度的思維轉型》，頁228-230, 234，臺北：元照。

[19] *See* KEVIN D. ASHLEY, ARTIFICIAL INTELLIGENCE AND LEGAL ANALYTICS: NEW TOOLS FOR LAW PRACTICE IN THE DIGITAL AGE 3-4 (1 ed. 2017).

doi:10.6342/NTU202103631

建立於可確信的規範性前提。雖然在演繹法中，正確的前提能夠擔保正確的結果，但演繹法的問題在於無法廣泛的適用於前提持續產生改變的事務領域中，且演繹法也沒有辦法彰顯事物與事物的關聯。[20]

除了演繹法本身的缺陷之外，相關研究以及實踐之所以沒有突破的第一個原因，在於現實中的法律並不是前提要件只要滿足，特定法律效果就一定會產生；在個案中會因為例外要件的滿足，而導致法律效果的不發生。面對例外無法事前窮盡的情形，應如何將程式結設計成能應對日常的運算公式則造成了困難。[21]第二個原因是法律本身與程式的差異在於，法律以文字為基礎，在法律規範的內涵上具有一定的不確定性，但程式是讓電腦理解的語言，內涵必須精確。另外，法律法律的基礎、來源眾多，且法源彼此之間有相異的位階、衝突關係必須解決。[22]在此理論以及實踐的障礙下，自動化決定的系統不僅花費許多時間加以建立，且沒有辦法如同人類法律適用者正確的作出決定。

這些困難隨著「資料為基礎」的自動化系統出現後逐漸消失。資料為基礎的自動化決定對於法律的自動化適用，提供了一種思維改變方式：**與其將個案輸入公式中找到答案，不如從個案之中找尋公式**（規律）[23]。以資料為基礎的人工智慧，透過分析眾多的法律文件（官方規範、裁判、契約、期刊），找出對於特定法律問題具有解決關聯性的文字與段落，再從其中分析、歸納出解決該問題的公式。機器學習在此時最大的助益就是取代過去人為編碼的過程，轉為自動的進行概念性的法律資訊提取（conceptual legal information retrieval），由電腦針對文件中的字串進行關聯性的比對，分析出對於個案關聯性高低程度不同的資料，並跨越過去因為資訊過多無法實踐的瓶頸，協助

[20] *See* ERIK J. LARSON, THE MYTH OF ARTIFICIAL INTELLIGENCE: WHY COMPUTERS CAN'T THINK THE WAY WE DO 107-115 (1 ed. 2021).

[21] *See* Giovanni Sartor, *Defeasibility in Legal Reasoning*, *in* INFORMATICS AND THE FOUNDATIONS OF LEGAL REASONING 119 (Zenon Bankowski, et al. eds., 1995).

[22] *See* Edwina L. Rissland, *Artificial Intelligence and Law: Stepping Stones to a Model of Legal Reasoning*, 99: 8 YALE L. J. 1957, 1961 (1990).

[23] *See* Branting, *supra* note 17, at 13-16.

doi:10.6342/NTU202103631

建構出法律論理的運算模型，解決法律問題。[24]這種實踐方式的突破，一舉的降低系統建置所花費的時間。

此外，利用此種方式建構的人工智慧，雖然在方法上不同於人類理解以及解決法律問題的方式，但經過驗證，這些自動化系統在結果上高度的與人類適用的結果相符。人工智慧，並不是透過理解問題的本身而尋找問題的答案，而是透過自然語言的處理（Natural Language Processing）方式，透過字串與字串之間的關聯性高低，加以決定問題的答案。[25]因此，與其說這是一種法律適用行為，不如說是「模仿」法律適用行為，以人類法律適用的產物作為樣本加以學習，而不在乎真實法律適用的「過程」[26]。

在歸類上，這種模擬過去樣本所生的推論方式屬於「歸納法（induction）」，透過分類所觀察到的樣本以獲取經驗，並且透過此經驗而建立假設的推論模型[27]。此種推論能夠成功的關鍵，在於事實能夠重複的發生，因此透過學習過去樣本而生成的模型能夠在未來中繼續適用。在巨量資料以及機器學習的呈現下，資料中的內涵會如實的加以顯現。一方面，由於資料的內容決定了模型的結果，因此資料中愈常出現特徵，對於結果的影響愈強烈；而在資料中較少出現或沒有出現的特徵，對於結果的影響愈薄弱，在總體上形成「頻率推定（frequency assumption）」的結果。[28]另一方面，由於是利用巨量數據所呈現的模型進行推論，模型對於事件的適應性增加，能夠有效的應對不同的情形，不論結果是截然區分的是否二元區分，或是連續的數值，自動化決定都能夠對於詳細的對於每一種類型的個案給予一個特定的結果，個案的形成「量身訂做的規則適用模型」。[29]

---

[24] ASHLEY, *supra* note 19, at 11-13.

[25] 關於此種技術，又可分為statistical techniques與selectional restriction，前者透過分析字串與字串在訓練資料中出現的關聯性作為基礎；後者則是利用語言知識的關係，事先的標示出字串與字串之間語言知識上的關聯性。*See* Dean Alderucci, *The Automation of Legal Reasoning: Customized AI Techniques for the Patent Field*, 58 DUQ. L. REV. 50, 56-60 (2020).

[26] *See* Peter K. Yu, *Artificial Intelligence, the Law-Machine Interface, and Fair Use Automation*, 72 ALA. L. REV. 187, 210-213 (2020).

[27] *See* LARSON, *supra* note 20, at 115-132.

[28] *See id*. at 150-154.

[29] Frank Fagen & Saul Levmore, *The Impact of Artificial Intelligence on Rules, Standards, and Judicial Discretion*, 93 S. CAL. L. REV. 1, 31 (2019).

doi:10.6342/NTU202103631

目前，比較常見的研究以及應用，是將人工智慧運用於建置判決結果的預測系統。以過去人類法官所作成的案件作為資料，透過特定的演算法，模仿特定法院的判決方式。比較有名的例子，像是愛沙尼亞即在訴訟標的低於歐元 8000 元的民事小額訴訟中，推行人工智慧法官的審理方式，由兩造上傳書狀後，交由人工智慧作出決定，當事人若不服判決結果，可以向人類法官上訴。[30]此外，亦有研究針對美國最高法院[31]、歐洲人權法院[32]所作成的判決進行分析，並且建構出預測特定法院判決結果的模型。[33]

3.2.2　洞見發現型的人工智慧

除了模仿的能力之外，人工智慧的另外一種能力是「發現資料之間關聯性」，提供人類在現實中加以使用。本文將這些人類無法觀察或理解的資料關聯性稱為「洞見」，因為這些資訊在現實中有助於政策的執行。

一般而言，個別決定的作成必須以可確信的知識為前提，而人工智慧的出現逐漸改變了這個基礎。科學知識的發現以及鞏固是透過觀察事物的變化，針對該變化提出假設，為了驗證該假設從事實踐，在不斷的試錯與修正的過程之中，建立理論以生產知識。這種知識是以「理論為基礎（theory-based）」的知識。自從資料利用的技術出現，此種以假設、理論為前提的知識生產模式受到揚棄，因為機器學習能夠透過演算法自主的從資料中習得關聯性，並在現實中運用此關聯性。此種資料之間的關聯性在現實中由於具有高度的應用性，因此對於行動的實踐者而言，將直接利用此種「洞見」，而省略掉傳統科學中「假設」以及「理論化」的階段。[34]

---

[30] Eric Niiler, Can AI Be a Fair Judge in Court? Estonia Thinks So, WIRED (Mar. 25, 2019), ogy.de/y68e (last visited: Oct. 19, 2021).

[31] Matthew Hutson, Artificial intelligence prevails at predicting Supreme Court decisions, SCIENCE (May 2, 2017), ogy.de/ibnt (last visited: Jun. 21, 2021)

[32] Masha Medvedeva, Michel Vols & Martijn Wieling, *Using Machine Learning to Predict Decisions of the European Court of Human Rights*, 28 ARTIFICIAL INTELLIGENCE AND LAW 237, 263 (2020).

[33] 不過，此處必須說明的是，相關研究中的「預測正確率」是以特定模型在「訓練資料」中的表現結果加以呈現，而這些模型在「現實」中的表現究竟如何，則是這些研究著墨不深的點。

[34] Chris Anderson, The End of Theory: The Data Deluge Makes the Scientific Method Obsolete, WIRED (Jun. 23, 2008), ogy.de/1gy7 (last visited: Jun. 16, 2021).

doi:10.6342/NTU202103631

機器學習以電腦的觀點所習得的「關聯性」，對人類世界而言究竟具有何種價值，在未經科學驗證之前皆無法確定。機器學習自資料中所習得的關聯性可能是人類世界所承認的固有知識的內涵之一；機器學習自資料中所習得的關聯性亦可能是人類現階段知識無法掌握的知識外的資訊，例如基因與特定疾病的關聯性；機器學習習得的關聯性可能是恰巧的關聯性，例如分辨狼與狗的模型是透過照片中是否有雪作為判斷標準；機器學習自資料中所習得的關聯性也可能是對於人類社會沒有用，甚至具有傷害的資訊，法學上著名的例子就是人工智慧會透過與法律所保障的特徵連結的「中介（proxy）」與決定目的的關聯性。以中介作為標準進行分類，在特定歷史、社會、文化脈絡下，將使表面上看起來無害的中介與法律上所禁止分類的特徵加以相連，產生事實上的歧視結果。例如：郵遞區號在表面上雖然為中性、不具有規範性禁止意涵的特徵，然而其在實際上相當於居住區域；而如居住區域有貧富的差別。例如：在美國貧富的差距，會與黑白的人種互相連結，因此在使用上郵遞區號就相當於種族的中介。[35]

傳統上，資料與模型都僅是發現知識的工具。而巨量資料的利用則挑起新一波的知識革命，認為模型以及資料本身即可能構成知識的本質。不少學者以「理論為基礎」的傳統知識發現方式加以辯護，例如：知識的發生尤其必須透過比較的觀察發現不同，釐清是什麼東西有所不同，單純資料的分析只能告訴我們有「不同發生」，但無法清楚的說明什麼不同以及為何不同[36]。另外，從「資料之間關聯性」的本質，就能夠說明這些洞見離真的「知識」還有一段距離。詳言之，「關聯性」並非「因果關係」。即便我們觀察到公雞叫後伴隨的現象是太陽升起，這並不代表公雞叫與太陽升起兩個事件之間具有因果關係。關聯性之所以無法代表因果關係，至少有以下三個原因[37]：一、因果關係是有原因與結果的方向關係，但關聯性僅描述了現象，無法說明事

[35] *See* Anya E.R. Prince & Daniel Schwarcz, *Proxy Discrimination in the Age of Artificial Intelligence and Big Data*, 105 IOWA L. REV. 1257, 1267-1283 (2020).

[36] Sabina Leonelli, Scientific Research and Big Data, THE STANFORD ENCYCLOPEDIA OF PHILOSOPHY (Summer 2020 Edition), Edward N. Zalta (ed.), ogy.de/zbca; Fulvio Mazzocchi, *Could Big Data be the End of Theory in Science?, A Few Remarks on the Epistemology of Data-driven Science*, 16 EMBO REP. 1250, 1250-1255 (2015).

[37] 以下的例子為王鵬翔教授與張永健教授於文章中所舉之例子。請參見王鵬翔、張永健（2015），〈經驗面向的規範意義—論實證研究在法學中的角色〉，《中研院法學期刊》，17期，頁252-256。

doi:10.6342/NTU202103631

件之間的方向，例如 X 事件導致 Y 結果的產生，但 X、Y 同時出現的關聯性描繪，無法說明「X→Y」的方向關係；二、兩個現象可能是基於另外一個原因而同時出現，彼此間可能根本沒有關係，例如氣壓計指針的數值上升與乾爽兩事件的同時發生，是因為氣壓本身的高低所影響；三、干擾因素的存在，導致錯誤的關聯性，例如統計結果顯示男性的平均薪資較女性高，結果是因為「年資」以及「教育程度」的干擾因素所導致的結果。換言之，這些關聯性還必須進一步進行分析才能夠作為因果推論的基礎。

不過資料關聯性的「洞見」經常能夠協助公部門執行任務，尤其能夠有效的部署不足的行政資源。詳言之，單一的公部門經常必須面對多數的人民或多數的案件，而行政資源的缺乏造成行政任務的資源分配不均，無法有效率的處理這些任務，而機器學習能夠有效的協助行政機關找尋資料之間的關聯性，協助公部門任務的履行。[38]舉例而言，在給付行政的情況中，在判斷申請福利給付的對象是否符合法律所規範的要件時，以機器學習區分申請者的情形，並且即時對於需要福利的對象核准給付[39]；在干預行政中，警察行政基於警力資源有限，透過機器學習區分個案的危險以及預測可能發生的風險，例如特定區域具有高度犯罪發生的可能性（犯罪熱點的辨識）[40]；或在金融交易的情況中，從股票市場的眾多交易中透過機器學習追緝可能的非法行為，從多數的交易資料之中觀察到非法交易（內線交易、證券詐欺）發生的模式，並且加以即時遏止[41]。

從以上的說明可以得知，人工智慧能夠從資料之中找到關聯性。這種資料間的關聯性雖然不一定是人類社會中既有的知識，但由於關聯性的存在，公部門即能夠透過此種關聯性的輔助，遂行履行公共任務等其它社會控制的目的。

***

[38] *See* DAVID FREEMAN ENGSTROM, et al., GOVERNMENT BY ALGORITHM: ARTIFICIAL INTELLIGENCE IN FEDERAL ADMINISTRATIVE AGENCIES 21-69 (2020), available at: https://ogy.de/b1kr.

[39] *Id*. at 39-41.

[40] *Id*. at 31-34.

[41] *Id*. at 23-25.

doi:10.6342/NTU202103631

本節說明了人工智慧的兩種能力，如何在法律領域中發展，以及如何進一步協助公部門進行法律適用的任務。在應用上，「模仿」與「洞見」這兩種能力，不僅有效的協助公部門進行法律適用結果的預測以及事實的認定，尚能在其它的面向上有所協助，例如：規範的分析、過濾大量的文件並提取出重要資訊、分析文件生成論證、草擬文件、趨勢分析、文件審閱…等能夠有效降低人力資源的資訊過濾工作。[42]

面對這樣的現實，我們此時應該繼續詢問的問題是：我們是否「應該」將法律適用的任務交給人工智慧來從事？從以上的分析，我們可以說人工智慧所發現的「洞見」，**不應該**直接用以從事法律適用的決定，理由在於以理性為基礎的法律適用任務，應以具有因果關係的經驗事實為基準，然而資料之間的關聯性並非因果關係。至於因為人類因為控制的慾望而去利用「洞見」，則是法律對於人類究責古典問題的借殼上市，諸如公部門利用這些洞見來進行決定，是否違反「恣意禁止原則」。這樣的問題，並未對於法律的本質帶來理論上的挑戰。相反的，對於法律適用而言，最應該釐清的問題應該是：是否應該讓「模仿型」的人工智慧進行法律適用的任務？

回答這個問題的關鍵，尤其必須分析人工智慧在適用法律時，是否對於法律適用的內涵，產生了重要的改變。[43]唯有發現這個改變，我們才能進一步的進行價值判斷，釐清這種改變到底是不是我們接受的改變。

[42] *See* Dana Remus & Frank Levy, *Can Robots Be Lawyers: Computers, Lawyers, and the Practice of Law*, 30 GEO. J. LEGAL ETHICS 501, 508-530 (2017).

[43] *See generally* NEIL POSTMAN, TECHNOPOLY: THE SURRENDER OF CULTURE TO TECHNOLOGY (1 ed. 1993).

doi:10.6342/NTU202103631

3.3　法律適用的本質

為了回應「是否應該讓模仿型人工智慧從事法律適用任務」的問題，本節將進一步勾勒法律適用的本質，以作為後續分析模仿型人工智慧對於法律適用的內涵產生何種挑戰。

3.3.1　法律論證作爲法律適用的方法

法律的適用，目的在於釐清個案事實是否在特定的規範範圍之內，以決定規範的法律效果是否會在具體的個案之中實現。然而，法律透過文字作為承載規範意涵的載體，文字本身的意涵不確定性，導致規範在面對到各種不同的案例事實時，產生意涵的不確定性；為了解決規範內涵的不確定性，法律適用者透過「解釋」的方式釐清個別法律的規範意涵。[44]所謂的「法律解釋」，主要透過「法律論證（legal reasoning; legal argumentation）」的方式加以進行，透過提出特定的理由，以支持法律在特定案件的情況中應為特定內涵的規範性主張。[45]

然而，由於法律本身具有規範意涵的不確定性，因此如何判斷支持特定解釋的論證是否正確，即成為了困難的問題。在眾多分析法律論證正確性的理論中，其中尤其以 Robert Alexy 所提出的「法律論證理論」為我國學界熟悉。[46]以下以 Alexy 的理論，作為討論法律適用本質的基礎。

一般性而言，法律的適用在方法上是採取三段論法（syllogism）的演繹模式，其步驟可以說明如下[47]：

大前提：假使任何一個案件事實實現 R，則應賦予其法效果 E（R→E）

小前提：特定案件事實 F 實現 R，質言之，其係 F 的一個「事例」（F=R）

---

[44] *See* Mireille Hildebrandt, *A Vision of Ambient Law*, *in* REGULATING TECHNOLOGIES: LEGAL FUTURES, REGULATORY FRAMES AND TECHNOLOGOCAL FIXES 180-185 (Roger Brownsword & Kren Yeung eds., 2008).

[45] 顏厥安（1998），〈法、理性與論證〉，收於：顏厥安編，《法與實踐理性》，頁105、140-141，臺北：允晨。

[46] Robert Alexy認為法律論證是特殊的言說過程，因此法律論證的正確性，除了必須符合一般言說論證的規則之外，也必須符合法律特有的論證要求。ROBERT ALEXY, A THEORY OF LEGAL ARGUEMENTATION (Ruth Adler & Neil MacCormick tans., Oxford University Press 1 ed. 1989) (1978). 國內關於Robert Alexy法律論證理論的說明請參見顏厥安，前揭註45。

[47] Karl Larenz（著），陳愛娥（譯）（2013），《法學方法論》，初版九刷，頁169，臺北：五南。

doi:10.6342/NTU202103631

結論：對 F 應賦予法律效果 E（F→E）

形式的來說，若存在一法律 R，該法律由 R_1、R_2…R_N 數要件加以構成，該法律有 E 的法律效果。法律適用者在面對一事實 F（由 F_1、F_2…F_N 等部分加以組成）是否應該產生 E 的法律效果，而進行法律適用時，必須就 F_1 是否屬於 R_1 的類型、F_2 是否屬於 R_2 的類型、…、F_N 是否屬於 R_N 的類型進行判斷，進而推論出個案是否產生 E 的法律效果。

如此的思維邏輯，分別是明確化所適用的規範、汰選出符合構成要件事實樣態的法律事實，並且將法律效果適用於特定的個案之中。法律適用者在此三個步驟之中，皆有不同的任務：在大前提的步驟之中，確認應適用的法律規範，並且釐清法律規範的內涵；在小前提中，自日常事實中「裁切」出涉及法律特徵的重要事實，並且進行法律觀點的評價；在結論之中，將抽象的法律效果在具體個案之中實現。

依據 Alexy 之見，法律適用是否正確，必須分別從「內部證立（internal justification）」與「外部證立（external justification）」兩個角度加以檢視。[48]外部證立的關注對象，則是大前提內涵確定的過程，即大前提的內涵是否正確、真實以及妥當；內部證立的觀察重點，是小前提涵攝的過程，即認定特定事實是否符合法律要件的過程，是否為有效的邏輯推論。必須指出的是，三段論法的思維並不是專屬於具有成文法典的大陸法系，即便是案例法的情形中，法律適用者仍舊必須從過往的案件中找出蘊含於其中的大前提、說明大前提在目前的個案之中應如何適用，最終將個案置入於大前提之中得出法律效果[49]。

三段論法過程中的關鍵，在於如何將一個抽象的規範連結到每一個具體個案的過程。這個過程的困難性在於，法律的用語以及文字通常並非使用能夠與日常生活事實能直接對應的「基本概念」。此種「案例事實的描述與法規範的構成要件之間存在有裂縫」的情形，法律的適用者必須透過法律的解釋方法更進一步的精確化法律概念，直到構成要件能夠與法律事實加以演繹

[48] Alexy, *supra* note 46, at 221; 張嘉尹（2019），〈憲法解釋作為法律續造—一個方法論的反思〉，《中原財經法學》，43期，頁10。

[49] *See* NEIL MACCORMICK, RHETORIC AND THE RULE OF LAW: A THEORY OF LEGAL REASONING 43-47 (1 ed. 2005).

doi:10.6342/NTU202103631

相連時，此時才該當涵攝過程的任務完美達成。[50]舉例而言，法律適用者必須作成 F 是否屬於 R 的判斷。由於 R 是抽象的法律文字，因此必須透過解釋的方式釐清 R 的要素，將 R 依據其意涵展開為 $R_1+R_2+\ldots+R_N$ 為而後論證 F 是否與 R 各內涵的要件相符。[51]因此，涵攝的成功仰賴透過解釋對於法律內涵加以釐清。

個案中的事實是否符合法律構成要件，有時候單純以感知為基礎就可以作成判斷，例如：人出生的時點、土地的位置，這些要件在判斷上並不會有問題[52]。但其它無法透過感知加以認定的要件，則需要透過解釋的方式釐清規範的內涵。Alexy 將法律論證之中外部證立的方式區分為以下五種：經驗論證（empirical reasoning）[53]、解釋要素（canons of interpretation）[54]、法釋義學論證（dogmatic reasoning）、裁判先例以及其它特殊的法律論證。[55]。

不過，以上的五種論證方式僅提供了個案的法律適用者，達成正確法律論證的手段；這些論證方式本身並未說明彼此之間具有何種關係。例如經驗論證之中，相關可以認定經驗的標準可能具有多種可能，但何種經驗應該被法律適用者加以援用，取決於法律適用者的認定，由各該法律適用者釐清何種經驗具有重要性；不同的解釋要素之中，除了以文義解釋作為解釋的基準之外，若有其他的論據，仍舊可以採取目的解釋等不同解釋的方法，偏離文義；而法釋義學解釋以及裁判先例的運用之中，也不禁止個案透過理由的陳述而偏離過去的見解。[56]因此，可以說這五種外部證立的方法，僅是外部論證中的「形式」，並非解釋的「方法」，也因此這些工具不具有次序上優先劣

[50] 王鵬翔（2005），〈論涵攝的邏輯結構—兼評Larenz的類型理論〉，《成大法學》，9期，頁13。

[51] ALEXY, *supra* note 46, at 222-230. 王鵬翔，前揭註50，頁14。以某甲判斷砍斷某乙右手之拇指、中指以及食指是否構成刑法第278條第1項中的「重傷」構成要件為例，王鵬翔教授指出必須依序透過以下的過程方能達成小前提中的涵攝任務：(1)凡使人受重傷者，應處五年以上十二年以下有期徒刑→(2)凡毀敗一肢之機能者，屬於重傷→(3)喪失手之作用，係毀敗一肢之機能→(4)如果手指失去效用，即喪失手之作用→(5)砍斷他人左手拇指、食指與中指，則手指即失去效用→(6)某甲砍斷左手他人拇指、食指與中指→(7)某甲應處五年以上十二年以下有期徒刑。

[52] 參見Karl Larenz，前揭註47，頁187-188。

[53] 經驗論證指的是某些構成要件的判斷必須藉助社會經驗，例如物品是否構成買賣法上的「瑕疵」，而對於經驗的內涵在個案中可能有多重標準，此時應訴諸相關跨領域的研究，方能確認經驗的內涵，以及個案之中應該使用的經驗，參見Karl Larenz，前揭註47，頁189-191。

[54] 即文義解釋、目的解釋、體系解釋、歷史解釋、合憲性解釋等的解釋方法。

[55] ALEXY, *supra* note 46, at 230-232.

[56] 顏厥安，前揭註45，頁163。

doi:10.6342/NTU202103631

後的關係，個案中究竟應採取何種論證方式，應視支持特定論證方式的「理由」具有何種重要性而定。[57]

從以上的說明可以得知，法律論證作為法律適用的方法，起於外部證立、終於內部證立。外部證立中對於法律規範內涵的解釋，必須視支持特定解釋方法背後的理由具有何種程度的重要性而定。在完成外部證立釐清規範內涵後，可以進一步從事內部證立，透過語意規則將抽象的規範與具體的事實加以連結。以上對於法律適用過程的說明，指出法律適用的過程必須遵守一些基本的程序，但這些程序僅在確保法律適用過程的「合理性」，並非預先的要求特定解釋結果。此處必須指出的是，法律適用合理性的確保，是透過法律適用者所給予的文字理由加以判斷，從文字的理由中判斷外部證立中採取特定解釋方式的理由是否充足、特定解釋方法的操作是否正確以及內部證立中的邏輯演繹是否無誤。合理的法律適用，不在於要求法律適用結果的一致，而是要求支持個案應特殊解釋的理由，必須具有普世性，在其它具有相同背景條件的情況中，亦能加以運用。[58]在操作上而言，隨著不同個案的累積，各種不同的理由能夠建築出一套縝密的規範適用體系，區辨各種理由的適用情形，穩定整體法律適用的過程、使法律適用具有可預測性。

有問題的是，若規範前提皆相同，法律以及法律適用的本身，是否有不同認定的可能？這涉及的是「法律可擊敗性」的爭議。

3.3.2 法律面對多變事實的方式

法律的可擊敗性指的是，特定規範即便要件皆滿足，其法律效果可能因為其它的原因，導致法律效果的不發生。[59]法律可擊敗性所涉及的問題是法律

[57] 不同的論證方式，背後都是基於不同的理由，特定的理由A可能支持文義解釋、特定的理由B可能支持目的解釋、特定的理由C可能支持歷史解釋...等。基於立法者能夠拘束行政（依法行政）以及司法（依法審判），特定個案中原則上須採取文義解釋；然若理由B、C的重要性高於A時，法律適用者可以透過給予理由，說明個案之中應採取目的解釋或歷史解釋的原因。針對理由之間重要性的比較，涉及權衡的問題，請參考本文3.3.2以下的內容。

[58] 法律的適用容許法律適用者提供正當的理由，區別各該案件的適用與否。由於法律在效力上仍舊必須具有普遍性，因此該正當的理由是個案的特徵在各種情況的評價都會相同的時候，即能構成個案偏離法律要件的原因。MACCORMICK, *supra* note 49, at 88-91.

[59] *See* H. L. A. Hart, *The Ascription of Responsibility and Rights*, 49 PROC. ARISTOT. SOC. 171, 175 (1948-1949).

doi:10.6342/NTU202103631

如何調節「規範性」以及「事實的複雜性與可變性」之間的衝突。[60]詳言之，一般性的法律規範，若要建立其規範性，法律則應該一視同仁的適用，不應具有例外；即便有例外，立法者也應該事前的將例外置於規範之中。在此觀點下，法律不應具有可擊敗性，此才能夠建立法律的規範性。然而，由於事實具有多變性，特定規範的背景條件可能因為時間的經過而有所不同、立法者沒有想到該種特殊的情形，或立法者根本錯誤作成評時，法律體系的本身是否有能力解決這種衝突即有疑問。

法學中「原則與規則的區分」可以有效的說明法律為何應該具有可擊敗性。[61]「原則與規則區分理論」認為法律中有兩種規範，分別是「原則（principle）」以及「規則（rule）」，兩者差異在於內涵以及運作的方式有所不同。

原則指的是特定有效的法秩序中作為基礎的結構，這些「結構」表達了特定法秩序的價值與目的，而具有「最佳化要求（或稱「盡力實現的誡命」optimization command）」的特性，亦即原則在適用上必須在事實上以及法律上盡可能的加以實踐[62、63]。原則在適用上僅指引個案特定的論證方向，法官必須在個案中討論該原則對於個案的影響，但原則本身無法決定性的對於個案

[60] *See* Juan Manuel Perez Bermejo, *Principles, Comflicts, and Defeats: An Approach from a Coherentist Theroy*, *in* THE LOGIC OF LEGAL REQUIREMENTS: ESSAYS ON DEFEASIBILITY 288 (Jordi Ferrer Beltrán & Giovanni Battista Ratti eds., 2012).

[61] 在理論的發展沿革上，原則與規則的區分最早是由Ronald Dworkin加以提出並用以攻擊法實證主義者H. L. A. Hart規則適用體系背後的法律與道德的「分離命題」，後由德國法學者Robert Alexy批判性繼受、建立起以原則為核心的憲法基本權理論，並由荷蘭邏輯學家Jape Haag加以修正說明理由對於規範的功能，其中Robert Alexy的理論尤其具有重要性。國內對於此三者理論的研究文獻參見：雷磊（2009），〈法律規範的同位階衝突及解決：以法律規則與法律原則的關係為出發點〉，《臺大法學論叢》，38卷4期，頁1-66。國內對於Robert Alexy基本權理論的研究，中文文獻可參見：顏厥安（1998），〈法與道德 - 由一個法哲學的核心問題檢討德國戰後法思想的發展〉，收於：顏厥安編，《法與理性實踐》，頁67-69，臺北：允晨叢刊；張嘉尹（2002），〈法律原則、法律體系與法概念論 - Robert Alexy法律原則理論初探〉，《輔仁法學》，24期，頁1-48；陳顯武（2005），〈論法學上規則與原則的區分 - 由非單調邏輯之觀點出發〉，《臺大法學論叢》，34卷1期，頁1-45；王鵬翔（2005），〈論基本權的規範結構〉，《臺大法學論叢》，34卷2期，頁1-60。

[62] ROBERT ALEXY, A THEORY OF CONSTITUTIONAL RIGHTS 47 (Julian Rivers trans., Oxford University Press 2002) (1994).

[63] Ronald Dworkin對於原則的定義是「正義、公正或其他道德方面的要求」。RONALD DWORKIN, TAKING RIGHTS SERIOUSLY 82 (1 ed. 1977). 不過「原則」的內涵以及整體的原則體系，本非Dworkin的發展重點，也因此Dworkin對於原則的內涵定義較不精確。

doi:10.6342/NTU202103631

作出終局性的結果。[64]單一原則對於個案之中會產生單一方向的指引；而一個個案之中會因為多數不同方向的原則同時產生，構成了原則之間的衝突。面對原則的衝突，法律適用者必須透過「權衡（balancing）」的方式，來判斷個案中哪一個原則在個案事實中，具有重要性，並且使具有重要性的原則指引個案最終的解釋方向[65]。在此情形中，法律適用者在個案中建立一「有條件的先後次序」，使個案依據這個先後次序適用原則，由具有優先次序的原則在該個案中勝出，對於個案的最終決定產生指引的效果。[66]需注意者是，由於原則具有「最佳化要求」，因此原則之間的位階先後次序關係僅在個案中有適用的可能，而不具有一般性，不同個案中原則之間的先後次序關係有所不同。就此而言，原則對於個案的效力僅具有初顯性（prima facie character）[67]，會因為權衡的結果而效力不繼續存在。因此，在效力上原則對於個案的理由僅是「促成性的理由（contributing reason）」[68]。

肇因於原則在適用上權衡結果的不確定以及對於個案作用方向的不確定[69]，立法者基於法明確性的要求必須先進一步加以考量個案的原則衝突並超前部署，事先的在各種案件之中綜合權衡各種同向、反向原則，而做出的立法決定形塑「法律規則」[70]，以有效解決法律適用不確定性，避免法律適用的無效率[71]。在此意義下，規則與個案的結合，取代了其他原則在個案中考量的需求，賦予個案「決定性的理由（decisive reason）」，終局性的決定了事實的規範性評價。[72]因此，規則在運作邏輯上，是以「全有全無（all-or-nothing）」的方式進行適用，只要個案的事實符合了規則的構成要件，法律效果則會直接產生。就像是棒球規則，打擊者獲得三個好球就要出局。[73]

---

[64] DWORKIN, *supra* note 63, at 25.
[65] *Id*. at 26.
[66] ALEXY, *supra* note 62, at 50-51.
[67] *Id*. at 57.
[68] Jaap Hage & Aleksander Peczenik, *Law, Moral and Defeasibility*, 13 RATIO JURIS 305, 307 (2000).
[69] *Id*. at 309.
[70] JAAP HAGE, REASONING WITH RULES: AN ESSAY ON LEGAL REASONING AND ITS UNDERLYING LOGIC 111 (1 ed. 1997).
[71] Hage & Peczenik, *supra* note 68, at 309.
[72] *Id*. at 306-307.
[73] DWORKIN, *supra* note 63, at 24-25.

doi:10.6342/NTU202103631

原則與規則之間的效力差異，意味著特定法律事實間若存在「規則」，個案之中即無適用「原則」的空間，而構成了規則的「排它性」。[74]這邊產生的問題是，規則「為何」具有排它性？這個排它性到底多強？是一個沒有例外的「強排它性」？還是有例外的「弱排它性」？[75]排它性的有無以及強弱，在面對到規則背後的理由與現實的狀態中不相符合、甚至錯誤的時候，會彰顯出來。換言之，排它性的承認，意味著「規則存在的事實」即賦予規則凌駕於原則的效力，而與規則的內容在現實中是否有正當性沒有關聯。[76]規則單純存在的事實就讓規則獲得效力上的排它性，一種可能的原因是規則的制定者在特定法秩序中的特殊地位。法實證主義者通常會從「效率」以及「權威」的觀點，說明規則的排它性是正當的，甚至證成其具有「強的排它性」。[77]由於本文目的不在為規則主義、或以規則為法律中心的法實證主義者辯護，因此無意在理論層次終局的解決這個問題。相反的，本文的目的是為現行實證法加以尋覓可以支撐的理論，亦即在一個落實依法行政、依法審判，同時賦予司法一定違憲審查權限的法治國家中，找到符合原則與規則之間合理互動的理論基礎。一個比較符合這種現實的理論版本，毋寧是承認規則具有排它性，但僅是一個弱的排它性。理由在於弱排它性在現實中的模樣，就是法律適用者（尤其是司法）原則上以規則的適用為優先，不因為規則的內容不具有正當性就不使用；僅在規則內容的正當性超出一定的界線後，方得在個案中適用原則。

此時，更進一步必須釐清的問題是，弱排它性在現實中應如何實踐？如何避免因為原則的適用，而架空了規則在法體系中的地位？設想在事實變遷的情況中，若 A 時刻所制定規則，其背後的理由與 B 時的狀態不相符合、甚至錯誤的時候，原則與規則兩者應如何互動呢？支持弱排它性者指出，此時並非直接捨棄規則而去適用原則，而是重新思考並衡量個案中彼此衝突原則

[74] 王鵬翔（2008），〈規則是法律推理的排它性理由嗎？〉，收於王鵬翔編：《2008法律思想與社會變遷》，頁348，臺北：中央研究院法律研究所。

[75] 關於這個問題，Joseph Raz與Robert Alexy間有相當大的爭論。簡單的做歸類，Raz是透過「權威論點」證成規則具有「強的排它性」；而Robert Alexy則認為規則僅具有「弱的排它性」。詳細討論，請參見王鵬翔，前揭註74文之討論。

[76] 法理學界稱此為「獨立於內容的理由」。請參見顏厥安（2002），〈規則、理性與法治〉，《臺大法學論叢》，31卷2期，頁18-19。

[77] 參見王鵬翔，前揭註74，頁360；顏厥安，前揭註76，頁19。

doi:10.6342/NTU202103631

之間的關係。[78]由於規則是立法者權衡多個原則後所作成的決定，因此，理論上所有的規則都可以還原成多數原則權衡的狀態。[79]若規則在制訂的時候，支持特定解釋方向的原則沒有被立法者納入考量、影響規則的背景條件有所變化或原則的權衡有錯誤詞，該規則的排它性就因此消失，不排除原則在個案中的適用可能性。法律適用者則應該重新就衝突的多數原則之間進行權衡，就特定規則必須繼續適用的「形式理由（formal reason）」（包括任何現行規則應遵守、現行規則所體現的價值等其它理由）與未事先考量的原則之間進行權衡，作成個案的權衡。[80]

在原則與規則之間的互動下，我們可以說規則的適用是具有可擊敗性的。由於規則的制訂，是立法者在特定時空背景下，考量有限因素所作成的立法決定，該立法決定僅在該時空背景下具有「相對的」正當性[81]；若時空背景不同，這個立法決定背後的理由即可能不再正當，而必須重新權衡，釐清新的時空背景之下，相互衝突原則之間的關係。[82]

以上關於原則與規則互動的說明，一方面維持了法律必須平等適用的要求，另一方面也指出了法律系統應對多變事實的方法。詳言之，基於平等的要求，法律的適用是全面性、一般性的，因此規則具有「排它性」，原則上避免個案的權衡；然而，法律生於社會，必須有能力應對社會中多樣化的事實，而「原則」的存在，以及規則僅具有「弱的」排它性的條件下，促成了法律本身的彈性，使法律系統能夠應對多變的事實。[83]對於法律適用者而言，原則與規則的存在，意味著法律適用者有義務判斷個案與規則的背景條件是否有差異、差異為何、差異在規範上是否有重要性，並且判斷規則的排它性是否存在。

---

[78] Hage & Peczenik, *supra* note 68, at 311.

[79] 王鵬翔，前揭註74，頁366；王鵬翔（2007），〈規則、原則與法律說理〉，《月旦法學教室》，53期，頁81。

[80] HAGE, *supra* note 70, at 112.

[81] *Id*. at 117.

[82] Jaap Hage, *Law and Defeasibility*, 11 AI & L. 221, 225 (2003).

[83] *See* Richard M. Re & Alicia Solow-Niederman, *Developing Artificially Intelligent Justice*, 22 STAN. TECH. L. REV. 242, 259-260 (2019).

doi:10.6342/NTU202103631

3.3.3　小結

在本節之中，本文首先說明法律適用因為文字不確定所導致解釋的必要。而外部證立中的各種解釋方法不預設法律解釋的結果，但在適用的過程之中，必須透過理由的提供以正當化解釋的內涵。此外，由於理由的正當性取決於是否具有普世性，在法律適用過程之中，具有普世性的理由逐漸建築了縝密的法律適用體系。不過，適用法律的過程是個案性的，必須就個案的特殊性，與過往的理由的前提加以區辨，達到個案的法律正確適用。此外，法律的解釋與適用必須注意到規範環境是否有所更動，規範環境的更動可能造成法律解釋的背景條件也有所變更。法律系統的本身，為了處理規範環境變動的前提，透過原則與規則的相互協力，促成規範內涵變動的可能性。因此，法律的解釋與適用的本身具有彈性，並不預設解釋的結果，而是盡可能的在個案中，在考量所有條件下，確保法律適用結果的正當。

doi:10.6342/NTU202103631

3.4 模仿型人工智慧對於法律適用的影響

在釐清法律適用的內涵後，本文要進一步分析是否應該讓模仿型的人工智慧從事法律適用的任務。此處必須說明的是，要分析人工智慧是否能有能力從事法律適用的能力，並非透過分辨人工智慧能否作成「正確的」決定加以釐清。除了何謂完全正確的法律適用決定，是一個很難說明清楚的問題外，人類也不是總能作成正確的決定。若以這種方式作為評估的基礎，毋寧是對於人工智慧過度嚴格。此外，我們也不應該以人工智慧可能作出錯誤、或歧視的決定，而直接得出反對使用人工智慧的結論，因為人類的法律適用決定也經常是錯誤的、歧視的。[84]相反的，我們應該釐清的問題是，模仿型人工智慧所進行的法律推論，到底是哪種推論方式？這種方式與人類從事法律適用的推論，在過程上、結果上究竟有何種「差異」？透過檢討這些差異在現行法律秩序下是否應該加以容許，才能判斷是否應該透過人工智慧從事法律適用的任務。

模仿型的人工智慧，在推論方式上使用的是「歸納法」，從樣本中釐清出蘊含於樣本背後的規律。歸納法的特色在於，忠實的將蘊含於樣本中的趨勢加以呈現。[85]這樣的特色雖然是歸納法的優點，但同時也造成了歸納法在實踐上的障礙，即「僅」能反應樣本中的內容以及趨勢。並未存在於樣本中的案例以及趨勢，就沒有辦法反應在歸納法的推論結果中。現行第二波人工智慧透過增加樣本的數量補足這個缺陷，以巨量資料作為樣本，盡可能的擴充人工智慧進行推論的基礎，以彌補這種推論方式本質的缺陷。而演算法在人工智慧中的角色，則是以超越人的能力，盡可能的找到樣本背後的規律，使基於巨量資料為樣本的歸納法，能夠在多變的現實中加以運用，使「模仿」的能力有可能在現實中實現。

直觀的來說，人工智慧的確能夠透過模仿人類進行法律適用任務的決定方式。依據本章 3.2.1 的說明，我們可以發現模仿型的人工智慧已經在法律適用的應用中，取得了一些正面評價。透過巨量資料的輔助，自動化決定雖然

[84] 參見Luís Greco（著）、鍾宏彬（譯）（2021），〈沒有法官責任的法官權力：為什麼不許有機器人法官〉，《月旦法學雜誌》，315期，頁177-179。

[85] *See* LARSON, *supra* note 20, 115-121.

doi:10.6342/NTU202103631

能夠透過大量的個案建立極度精細的裁量規則，使每一種案例類型都有對應的法律效果[86]，模仿過往各種案例中的適用結果，依據不同的行為樣態，對應不同的法律適用決定結果。

以上對於人工智慧於法律適用任務的正面評價，主要是對人工智慧就特定案例的「結果」進行評價，因為特定案例的結果符合過往的案例，因此認定特定人工智慧在表現上是良好的。然而，學者近期指出了模仿型人工智慧的能力缺陷，諸如運作過程無法理解、在語言意涵理解上的效能不彰、及無法打破成規[87]。此處將釐清這三個特色是否會影響法律適用的內涵。

3.4.1　無法提供法律適用的理由？

學說上對於人工智慧應用的一個批評來自於人工智慧的運作過程是一個無法理解的「黑箱」。「黑箱」這個概念，在自動化決定上，主要在形容一個決定系統在運作上，知道輸入為何、經過系統後的輸出為何，但無法確切的知道特定輸入「為何（why）」導致特定的結果。[88]與人類所作成的法律決定會「給予理由」有所不同，不乏學者指出公部門若利用人工智慧進行法律適用，將造成無法知道決定作成的原因的結果，影響、甚至剝奪人民請求司法合理救濟的可能性。[89]這樣的主張是不是合理，必須對於人工智慧的「黑箱」特色進一步分析。

事實上，任何新技術剛產生的當下，皆會因為無法理解技術本質，而產生這種質疑。此外，我們目前所廣泛使用的眾多技術，我們也不一定對於其

---

[86] *See* Anthony Casey & Anthony Niblett, *The Death of Rules and Standards*, 92 IND. L. J. 1401, 1407-1412 (2015).

[87] 邱文聰，前揭註16，作者手稿，頁2-7。

[88] *See* FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION 3 (1 ed. 2016).

[89] *See e.g.,* Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORDHAM L. REV. 1085, 1089-1098 (2018); Steven M. Appel & Cary Coglianese, *Algorithmic Governance and Administrative Law*, *in* THE CAMBRIDGE HANDBOOK OF THE LAW OF ALGORITHMS 166-167 (Woodrow Barfield ed. 2020); Emre Bayamlıoğlu, *Contesting Automated Decisions*, 4 EUR. DATA PROT. L. REV. 433, 442 (2018); Emily Berman, *A Government of Laws and not of Machines*, 87 B.C. L. REV. 1277, 1348 (2018); Kiel Brennan-Marquez, *"Plausible Cause": Explanatory Standards in the Age of Powerful Machines*, 70: 4 VAND. L. REV. 1249, 1288-1289 (2017); Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, 105 GEO. L. J. 1147, 1207-1209 (2017); Karen Yeung, *Why Worry about Decision-Making by Machine?*, *in* ALGORITHMIC REGULATION 28-29 (Karen Yeung & Martin Lodge eds., 2019).

doi:10.6342/NTU202103631

內涵有充分的理解。例如：大部分人無法理解基因的內涵，但目前我們在訴訟中能夠透過基因的檢測結果作為訴訟之中的證據；大部分人無法理解電子電路的內涵，但道路上卻廣泛的使用自動化號誌作為交通導引的機制；此外，對於我們自己「腦袋」的運作方式，即便腦神經科學至今仍未對於運作方式提出合理的科學說明，但我們仍舊透過人腦加以決策。

排除科技使用者基於自我保護目的、透過法律手段對於科技內涵所為的有意識的隱藏[90]所導致的法律上的黑箱（legal secrecy）[91]現象，過去使用科技所導致的無法理解，大部分主要指涉的是一般人不具有專業知識，而無法理解相關科技內涵的情形[92]。這種基於專業知識所生的知識落差，並不構成使用科技正當性的問題，理由在於相關科技內涵以及所產生的效果能夠透過專業人士加以把關，而不發生預期外的結果。

相較於過往科技因為專業知識所造成的理解性落差，機器學習的「黑箱」特色在於，不論是外行人或者內行人皆無法理解機器學習獲致結果的運作過程[93]。機器學習運作的過程之所以無法理解，在於機器學習的運作過程超乎人類能夠想像的變項數量。以人類的理解能力為基礎，人類可以想像兩個變項（二維）、三個變項（三維）之間互相的運作關係；但在四個變項或者四個變項以上的情況之中，資料之間的關係已經並非人類能夠透過思維模擬出來的線性（linear）、單調性（monotonicity）運作模式，而是非連續、不單調的運作模式。由於人類無法理解各種變項如何影響最終的結果，因此形成了人類無法理解的運作模式。[94]

從以上的說明，我們可以得知人工智慧的黑箱，指的是人類無法理解人工智慧作成特定結果的「原因（cause）」。然而，在法律論證的任務上，從來都不要求決定者給予決定的「原因」，而是要求決定者提出能夠正當化個案決定的「理由（reason）」。法律唯實主義者（legal realism）過去探討的

---

[90] *See* Jenna Burrell, *How the machine 'thinks': Understanding opacity in machine learning algorithms*, 3: 1 BIG DATA & SOC. 1, 3-4 (2016).

[91] 例如僅僅三行的資訊請求，卻以三百頁的資訊加以掩蓋，使人無法知悉確切的重要資訊為何，*see* PASQUALE, *supra* note 88, at 6.

[92] *See* Burrell, *supra* note 90, at 4.

[93] *See id*. at 4-5.

[94] *See* Selbst & Barocas, *supra* note 89, at 1094-1096.

doi:10.6342/NTU202103631

一個有趣問題是：究竟什麼影響了法官的判決結果？而最終的研究結論是，作成判決之前是否進食，才是影響個案的「原因」。[95]尚且不論這個研究是否正確，我們可以確定的是，即便我們無法理解「人類」本身的運作過程，現行法律仍舊能夠對於人類法律適用者進行課責。當代對於人類法律適用者課責，是透過「理由提供（reason-giving）」的方式，將該決定涵蓋於一個較具有普遍性的原則之中，以供第三者對於理由賦予的過程進行驗證，以提升決定的正當性以及可課責性。[96]以拘束我國公部門的規範為例，不論是司法行為[97]或者行政行為[98]，法律皆課予決定者提供理由的義務；適用法律的人類，必須在經過法制所要求的程序下提供論理，說明個案形成的過程，使第三人能夠理解其決定的理由，並且進一步的驗證其決定形成的過程是否正當。

換言之，黑箱所造成的決定原因無法理解，並不會影響是否能夠從事法律論證的原因。相反的，如果人工智慧能夠透過模仿，達到「給予理由」的程度時，人工智慧在法律論證中「給予理由」的要求上，也與人類沒有什麼太大的差異。這樣的觀點，與人工智慧之父雖然，智慧並非模仿，但給予理由作為一個形式的要求，並不以具有智慧為前提。而這樣的結果，與現行法律適用的實務中，也沒有太大的差異。現行適用法律的實務中，可以發現多數的時候，行政機關或者是法官皆是援用例稿，給了一個「心口不一」的決定。然而，這樣的決定，也並不直接違反任何實定法律規定。

---

[95] *See* Shai Danziger, Jonathan Levav & Liora Avnaim-Pesso, *Extraneous Factors in Judicial Decisions*, 108 P. NAT'L ACAD. SCI. USA 6889 (2011).

[96] *See* Frederick Schauer, *Giving Reasons*, 47 STAN. L. REV. 633, 641 (1995); Ashley Deeks, *Secret Reason-Giving*, 129 YALE L. J. 612, 629-34 (2020).

[97] 我國立法者課予民事法院、刑事法院、行政法院、司法院大法官明確化理由的義務。參見民事訴訟法第226條第1項第6款：「判決，應作判決書，記載下列各款事項：…六、理由。」、刑事訴訟法第223條：「判決應敘述理由，得為抗告或駁回聲明之裁定亦同。」、行政訴訟法第209條第1項第7款：「判決應作判決書記載下列各款事項：七、理由。」、憲法訴訟法第33條第1項第8款：「判決應作判決書，記載下列事項：…八、理由。」

[98] 行政機關作成的單方決定即為行政處分，行政程序法即要求行政機關有明確化作成理由的義務，就此可參行政程序法第96條第1項第2款：「行政處分以書面為之者，應記載下列事項：…二、主旨、事實、理由及其法令依據。」然而此理由課予的義務並非毫無例外，如行政程序法第97條即規範六種無須負擔理由義務的情形「書面之行政處分有下列各款情形之一者，得不記明理由：一、未限制人民之權益者。二、處分相對人或利害關係人無待處分機關之說明已知悉或可知悉作成處分之理由者。三、大量作成之同種類行政處分或以自動機器作成之行政處分依其狀況無須說明理由者。四、一般處分經公告或刊登政府公報或新聞紙者。五、有關專門知識、技能或資格所為之考試、檢定或鑑定等程序。六、依法律規定無須記明理由者。」

doi:10.6342/NTU202103631

若從這個觀點來討論這個問題，關鍵就在於人工智慧能不能給出一個理由。[99]對於這個問題，我們可以借用人工智慧之父 Alan Turing 提出的「圖靈測試」來加以檢驗。Alan Turing 對於電腦是否有「智慧」的判準，以電腦是否能夠通過「圖靈測試」。圖靈測試的內涵是「模仿遊戲（imitation game）」。這個想像的實驗中有三個角色，兩個人類跟一個電腦。其中一個人類 A，會在不知道對象的情況，分別以打字的方式，問另一個人 B 與電腦問題。A 必須從 B 與電腦所回覆的內容中，判斷誰是人類、誰是電腦。如果 A 沒有辦法清楚的分辨誰是電腦、誰是人類時，該電腦則通過了圖靈測試，而具有智慧。[100]筆者認為圖靈測試的問題在於，將「模仿」與「智慧」混淆，通過圖靈測試只能說具有模仿的能力，而並非具有智慧。但這個混淆不妨礙我們借用「圖靈測試」來回答此處的問題，因為我們要釐清的正是人工智慧是否具有模仿人類語言的能力。若人工智慧能夠通過圖靈測試，人工智慧當然也能從事法律適用的理由給予任務。就現在當下的科技現實來說，目前尚無完全能通過圖靈測試的人工智慧，但隨著相關科技的進步，愈來愈多科學家宣稱人工智慧通過圖靈測試只是時間的問題。[101]但理論上而言，人工智慧這種透過連結方式的歸納系統，是否有可能從事語言溝通行為？這個問題的答案，更進一步影響的是人工智慧是否同時能夠從事「內部論證」的問題。

3.4.2 無法針對新事物進行法律適用

本文於 3.3.1 指出，法律的論證，始於外部證立，終於內部證立，透過外部證立釐清規範的內涵後，透過語意規則，將特定抽象的法學概念有邏輯的推演到具體的事物，將法律概念與事實加以連結。換言之，人工智慧是否能夠發展出語言的能力，直接影響的是法律適用任務中的內部證立。

如前所述，在自然語言處理的領域中，人工智慧並非真的在理解問題的內容後，就問題進行回答，而是透過歸納法的原理，達到類似人類回答的結

[99] 不少學者即指出如果人工智慧能夠給出理由，沒有理由不去使用可以做的更正確的人工智慧，請參見Luís Greco，前揭註84，頁179-182；Eugene Volokh, *Chief Justice Robots*, 68 DUKE L.J. 1135 (2019).

[100] See Allan M. Turing, *Computing Machinery and Intelligence*, 59: 236 MIND 433, 433-434 (1950). Turing的問題在於錯把模仿的能力當作智慧。

[101] Evgeniya Panova, Which AI has come closest to passing the Turing test?, DATACONOMY (Mar. 9, 2021), ogy.de/jf8u (last visited: Oct. 19, 2021).

doi:10.6342/NTU202103631

果。在現階段，這種模仿人類回答的方式，在很基本的語言問題上會沒有辦法運作。例如在「*主管機關拒絕給予示威者遊行許可，因為**其**畏懼暴力*。」這個句子中，人工智慧無法回答「其」這個代名詞，指的究竟是「主管機關」，還是「示威者」。[102]對於人工智慧在語言能力的捉襟見肘，人工智慧學者提出的解方式混用第一波人工智慧與第二波人工智慧，透過以「規則為基礎」的人工智慧，將特定事物之間的關係以程式加以明確化，以解決第二波人工智慧在文義理解以及回答上的障礙。[103]

人工智慧在語言處理的脈絡中效能不彰，部分學者將原因歸結於人類的語言中，蘊含了很多未指明於文字中的一般經驗（common sense）以及默會知識（tacit knowledge）[104]，且語言的脈絡、語境會影響文義的理解。[105]這些一般經驗、脈絡以及語境，表彰的是事物在社會中的關聯。由於資料庫中缺乏這些經驗、知識等相關資料，因此人工智慧無法有效的從事語言溝通的任務。

本文認為以上的論點，是用「人類」進行語言活動的觀點，對於人工智慧進行資料處理的方式進行評價，而誤解的指出人工智慧沒有辦法進行語言活動的原因。理由在於，人工智慧本就不是透過「理解」的方式，來從事任務。因此即便將一般經驗刻意編碼放入資料庫中，人工智慧也不會因此就能夠補足其在語言能力上的缺陷。核心的問題應該是人工智慧透過資料連結的方式，有沒有辦法學會這些默會知識。[106]技術的進展，讓我們發現到默會知識

---

[102] LARSON, *supra* note 20, 196. 書中的原句是：The town councilors refused to give the angry demonstrators a permit because **they** feared violence.

[103] 許聞廉（2017），具深度理解之對話系統及智慧型輔助學習機器人，科技部MOST 107-2634-F-001-005計畫，轉引自邱文聰，前揭註16，頁4。但第一波人工智慧所生的問題（參見本文1.3.3.2的內容）也會隨之產生。

[104] 依據哲學家Michael Polanyi的區分，知識可以分為可以用文字加以表達的外顯知識（explicit knowledge），以及具有不可表達（inexpressibility）、不可明確化（unspecifiability）的默會知識。後來Harry Collins將默會知識區分為關係型的默會知識（relational tacit knowledge）、身體性的默會知識（somatic tacit knowledge）以及集體社會性的默會知識（collective tacit knowledge）。參見邱文聰，前揭註16，〈第二波人工智慧知識學習與生產對法學的挑戰—資訊、科技與社會研究及法學的對話〉，頁148-149。

[105] *See* Harry Surden, *Artificial Intelligence and Law: An Overview*, 35 GEO. ST. U. L. REV. 1305, 1325 (2019); Harry Surden, *Machine Learning and Law*, 89 WASH. L. REV. 87, 105-106 (2014); Alderucci, *supra* note 25, at 60-69.

[106] 針對這個問題，研究現象學與存在主義的哲學家Hubert Dreyfus與研究人類如何取得專業知識的知識社會學者Harry Collins針對這個問題有一個辯論。詳細的討論請參見邱文聰，前揭註16，〈第二波人工智慧知

doi:10.6342/NTU202103631

並非不能由第二波人工智慧加以學習。例如，過去在醫療領域中，醫生對於醫療影響的判讀方式，是一種無法用言語描繪的默會知識，然而在現今人工智慧的實踐中，我們發現人工智慧能夠學習到辨識病灶的能力。[107]因此，是否能學習到默會知識的條件，僅在於充作人類經驗的訓練資料是否足夠、全面（holistic），而能夠反應現實。[108]就此，我們可以將人工智慧在語言能力上的實踐可能性作出一個推論：只要代表特定專業知識的語言樣本足夠能反應特定專業知識領域的全貌時，人工智慧在理論上就有能力在特定領域中，成功模仿人類的語言行為。

回到法學領域中，人工智慧在理論上是有能力從事模仿法律人的論證語言。理由在於法律人在論證所使用的論證語言，相對來說，是足夠全面的。法律人在從事法律適用的的過程中，若生解釋的需求時，典型的方式是去找尋過往類似案件的操作方式，例如司法判決、行政函釋，若無法解決，則去翻閱教科書、期刊、專書等學者對於相關爭議的見解，以當代法律社群所接受知識，於個案中加以援用，達成法律適用的目的。[109]換言之，若能將法律人在進行法律適用的這些知識，全面性的轉為訓練資料時，模仿型的人工智慧，在理論上也有能力模仿人類從事法律論證。

當然，以上的推論是建立在所有案例都會重複，而不會有新案例發生的前提。若有「新案例」發生時，而這些新案例並未事先被法律社群加以討論時，法律適用的人工智慧在此時即發生實踐上的困境。理由在於，模仿型法律人工智慧所依循的規範內容，是固定的、封閉的。因此，在面臨到過去沒有處理過的新事物時，人工智慧欠缺新事物在樣本中定位的訓練資料，因此沒有辦法透過資料之間的連結，對於這些新事物進行模仿以及法律論證。[110]

識學習與生產對法學的挑戰—資訊、科技與社會研究及法學的對話〉，頁144-155。

[107] 陳偉婷，〈台灣醫療科技新突破 人工智慧3分鐘揪肺癌病灶〉，《中央通訊社》，2021年2月26日，ogy.de/w72b（最後瀏覽日：2021年10月20日）。

[108] 參見邱文聰，前揭註16，〈第二波人工智慧知識學習與生產對法學的挑戰—資訊、科技與社會研究及法學的對話〉，頁144-145、155。

[109] 參見張嘉尹，前揭註48，頁7；邱文聰，前揭註16，〈第二波人工智慧知識學習與生產對法學的挑戰—資訊、科技與社會研究及法學的對話〉，頁159-160。

[110] *See* Simon Deakin & Christopher Markou, *Ex Machina Lex: The Limits of Legal Computability*, *in* Is Law Computable? Critical Perspectives on Law and Artificial Intelligence 55-57 (Christopher Markou & Simon Deakin eds., 2020).

doi:10.6342/NTU202103631

以上的論述，僅能推導出人工智慧在「新事物」中的運用有其侷限，而在重複發生的事物中，或許仍舊能仰賴過往的案例，透過類似判例遵循的原理進行法律適用。基於這個結論，人類與人工智慧在法律適用的差異，就僅在是否能夠對於「新發生的案例」進行法律適用，而重複發生的案例，即便使用人工智慧進行法律適用，也與人類沒有差異。但這樣的推論，忽略了法律解釋本身是有變化可能性，甚至可能必須典範變遷。法律在適用層次的內涵變遷，涉及到另外一個法律的特色—法律的續造。

3.4.3　無法進行法律的續造

廣義的來說，法律續造指的是法律適用者對於規範制定者沒有提供答案的法律問題，進行回答的一個動作。在這樣的理解之下，法律續造這個概念所包括行為，諸如：首度對於一個法律的內涵進行詮釋、偏離過往的意涵、在文義範圍內的規範意涵限縮以及規範範圍內的意涵皆屬之。在這種廣義法律續造的理解下，法律的續造與法律的解釋並非截然不同的行為，僅是一個過程的不同階段。

之所以需要法律續造，可能的原因包括立法技術不良、法制素養欠佳、社會變遷、情事更易，導致法律適用者主動的介入，來調整規範內涵。[111]就方法上來說，法律續造是一個「目的論」的思考，認為現行規範文義的內容，與立法目的或特定法律秩序下的目的並不相符，因此由法律適用者主動的來填補這些漏洞。[112]依據目的的來源，可以分為「法律內的法律續造」與「法律外的法律續造」。前者旨在填補立法目的與規範文義間的漏洞，即法律續造的結果將超越法律文義，但仍在立法者目的的範圍內。相關填補的方式，諸如目的性限縮、類推適用。後者則是以特定的原則作為基礎，詮釋的結果將超出立法者的目的，但仍在整體法秩序的基本原則之內。[113]

[111] 參見李建良（2018），〈法學方法與基本權解釋方法導論〉，《人文及社會科學集刊》，30卷2期，頁256。

[112] 參見Karl Larenz，前揭註47，頁277-279。

[113] 參見Karl Larenz，前揭註47，頁278。

doi:10.6342/NTU202103631

法律續造在表面上，呈現的結果雖然是現在的解釋與過往的解釋有所不同，但其蘊含的意義是：**支持偏離規範文義的理由，其重要性已經高於維持規範文義的理由**。詳言之，支持既往文義解釋是基於 A 目的所生的理由，而法律續造則是基於 B 目的所生的理由。若要在個案中進行法律續造採取新解釋，意味者 B 目的的重要性必須高於 A 原則。換言之，法律續造的過程，是不同理由之間，互相權衡的過程以及結果，而連結到本文關於法律可擊敗性或法律論證可擊敗性的說明。[114]而上述「法律**內**的法律續造」與「法律**外**的法律續造」類型的區分，其分類實益僅在於支持法律續造的B目的，在重要性上必須需高於支持維持規範文義解釋 A 目的的程度不同。

如果我們從法律續造的角度來觀察人工智慧，可以發現人工智慧沒有辦法從事任何程度的法律續造。[115]理由除了人工智慧沒有主動從事「權衡」的能力之外，另外一個理由是根本不存在能夠作為法律續造的資料樣本。

依據前述 3.3.2 的說明，不同的原則會在個案之中相互衝突。法律適用者應考量到衝突原則之間的犧牲與實踐的關係，透過「權衡」的方式，於個案中說明衝突原則之間適用的優先次序，並進一步解決個案的適用。就此以觀，我們可以得知權衡是一個「個案」、「例外」且富具「創意」的行為。[116]而人工智慧所仰賴的歸納法，僅能夠呈現總體趨勢，無法判斷哪一個原則，在個案的適用中具有重要性。

此外，人工智慧之所以沒有辦法從事對低程度的解釋偏離的原因在於，人工智慧是以訓練資料中的人類知識作為決定的基礎。詳言之，即便在法律領域中，現實中所需要的法學知識或許與法律文件有高度的重疊，但法律的續造，經常會發生在法律適用者，過去從來沒有面對過的事實（即自始的漏

---

114 參見王鵬翔（2004），〈目的性限縮之論證結構〉，《月旦民商法雜誌》，4期，頁27-29（說明目的性限縮的許可性）。

115 相反見解請參見Luís Greco，前揭註84，頁176。Greco認為人工智慧能夠進行法律續造，理由在於好的法律續造，僅是將已經被法律秩序承認的命題推導到其它事物之上。筆者認為Greco忽略了典範變遷在法律適用領域中存在的可能性，僅將法律續造限於比較小的法律內的法律續造。

116 此處必須說明的是，學習如何「權衡」，以及學習「原則」，是兩件事情。舉例來說，我們可以將大法官解釋過往的案件作為訓練資料，訓練出一個理解現行憲法原則的人工智慧，但這僅代表該人工智慧能以現行的原則進行決定。但學習了憲法原則的人工智慧，仍舊不會權衡。因為權衡並非依據歷來的多數見解所作出的決定，而是針對兩個原則之中，進行價值高低的比較以及取捨。

doi:10.6342/NTU202103631

洞），或因為立法背景條件改變所產生的規範疑義（即嗣後的漏洞）。在現實之中，充當國家法律適用者的人類，即便遇到有史以來的第一個案子、沒有處理過的案子、特殊的案子，或是學者沒有研究過的案子，法律適用者也不得拒絕適用法律作出決定，而必須回歸到最初的法律適用方法，「將目光往返於規範於事實之間」，對於規範的內涵作出澄清。

但這種情況若發生在人工智慧時，若指引特定結果的權衡資訊，並不在訓練資料之中，人工智慧即沒有能力作出適切的權衡、從事法律的續造[117]。換言之，人工智慧這種以人類知識為基礎的歸納推論方式，在面對到知識的不足時，將使法律的適用者無法針對事實變化、個案特殊性，而偏離既往的解釋。[118]這樣的結果，意味著人工智慧在法律適用上缺乏變化以及彈性，乃至於因應經驗事實的差異，而進行典範的變遷。

綜上所述，本文認為模仿型人工智慧與人類適用法律間，最重要的差異即在於，人工智慧沒有辦法從事法律的續造。現行的法律，作為一種制度，能夠與持續變動的社會之間彼此互動，而支撐此機制的其中一環就是原則的適用：面對個案，首先判斷其與過往的案例以及過往的規則是否相容，若不相容，如何運用原則作出調整。[119]就此而言，目前以資料為基礎的人工智慧，沒有辦法面對社會的變化、多元性達到典範變遷的可能。

---

117 *See* Surden, *supra* note 105, at 1323-1324.

118 Rónán Kennedy, *The Rule of Law and Algorithmic Governance*, *in* THE CAMBRIDGE HANDBOOK OF THE LAW OF ALGORITHMS 226-227 (Woodrow Barfield ed. 2020). 在傳統德國行政法學中「不確定法律概念的判斷餘地」與「裁量」分別屬於法律規範中立法者賦予行政機關的認定自主空間。德國法對於是否能夠透過自動化系統作成行政處分，於德國聯邦行政程序法（Verwaltungsverfahrensgesetz, VwvfG）在2015年修正的時候，將「以全然自動化作成之行政處分（Vollständig automatisierter Erlass eines Verwaltungsaktes）」納入法律的規範之中，增定了第35a條。在內容上，其限縮了「行政處分作成方式自由原則」的空間，要求此種行政處分必須存在有法規命令明文容許外，行政機關對於行政處分的作成無判斷餘地或裁量空間的時候方得為之。此種保有行政權個案裁量空間不被機器侵奪的見解與「規則與標準」的選擇有異曲同工之妙。Christian Djeffal, *Artificial Intelligence and Public Governance: Normative Guidelines for Artificial Intelligence in Government and Public Administration*, *in* REGULATING ARTIFICIAL INTELLIGENCE 285-286 (T. Wischmeyer & T. Rademacher eds., 2020). 中文文獻對於德國行政程序法第35條a的討論請參見謝碩駿（2021），〈論全自動作成之行政處分〉，收於：黃丞儀編，《2017行政管制與行政爭訟》，頁1-86，臺北：中央研究院法律研究所。先期出版：http://publication.iias.sinica.edu.tw/92015091.pdf

119 *See* e.g., Frank A. Pasquale, *A Rule of Persons, Not Machines: The Limits of Legal Automation*, 87 GEO. WASH. L. REV. 1, 45-54 (2019); Rebecca Crootof, *"Cyborg Justice" and the Risk of Technological-Legal Lock-In*, 119 COLUM. L. REV. F. 233, 246 (2019); Tim Wu, *Will Artificial Intelligence Eat the Law?: The Rise of Hybrid Social-Ordering Systems*, 119 COLUM. L. REV. 2001 (2019); Andrew C. Michaels, Artificial Intelligence, *Legal Change, and Separation of Powers*, 88 U. CIN. L. REV. 1083 (2020); Deakin & Markou, *supra* note 110, at 60-63.

doi:10.6342/NTU202103631

***

本章說明了人類與人工智慧在適用法律上的差異。本章首先說明人工智慧的本質為資料、演算法以及人為的影響。再者，本文指出人工智慧的兩種能力，「模仿」以及「發現洞見」。為探討人工智慧對於法律所生的根本衝擊，本文著眼於模仿型的法律適用人工智慧，分析其與人類在從事法律適用上的差異。本文首先勾勒了法學方法論上在處理法律適用的內容。本文指出，法律的適用至少必須包括三個內容，透過給予理由說明法律適用的過程、以語意規則從事符合邏輯演繹的內部證立，並且對於規範的內涵以及正當性進行外部證立。本文尤其說明了，外部證立中規範，在特殊情形中可能發生可擊敗，而在個案中不發生規範效果或者規範效果改變的情形。藉由以上對於法律適用內涵的描述性說明，本文進一步分析人類適用法律與人工智慧適用法律間的差異。本文認為，第二波的人工智慧理論上有能力能夠給予法律適用的理由以及進行法律的內部論證，但沒有能力對於前所未見的新事實進行法律適用以及檢討大前提的內涵從事法律續造。

本章的結論，回應了第二章的提問，指出人類與人工智慧在法律適用上的確有差異，而差異即在於是否具有法律續造的能力。然而，這個差異是否足夠重要到必須要求公部門維持以人類為中心的決定方式？這是本文於下一章要釐清的問題。

doi:10.6342/NTU202103631

第四章

人爲介入作爲公部門正當使用人工智慧的必要條件

本文於第二章說明了歐盟法中「人為介入」的管制方式，第三章說明了人類與人工智慧在適用法律、作成決定上的差異。第三章的論述一方面釐清了歐盟法背後規範意涵的正當性，另一方面提出了更進一步的問題：此差異在公部門中應如何看待，是否必然需要透過建立「人為介入」的制度加以管制？在此問題之下，本章的研究目的在於分析公部門的憲法框架如何評價前述人工智慧對於法律適用過程的改變，以及說明「人為介入」的管制措施，是否為能夠符合憲法要求的法制框架。

基於人工智慧與人類在進行法律適用最大的差異在於人工智慧無法從事法律續造。本章將於 4.1 中釐清法律續造任務在法治原則中的地位，釐清使用無法從事法律續造的人工智慧是否與法治原則有違。本章將指出，法律雖然為普遍適用的規範，但仍舊會發生涵蓋過廣或涵蓋過窄的情形，而有由法律適用者依據個案進行法律續造的需求。在以憲法作為法源基礎的國家中，可以說法律的續造是法治原則中的必要要求。而法律續造並未對於法治原則產生衝擊，而是將法治原則的內涵強化於法律適用過程的理由交代。

在釐清法律續造在法治原則中的地位後，本文進一步檢視透過人工智慧從事法律適用的任務，是否有無庸進行法律續造的正當原因（4.2）。本文分別檢討完美性論點、專業性論點以及法治原則應加以典範變遷的論點。本文以人工智慧的運作現實為基礎，指出人工智慧並非完美的決定方式；公部門即便因為權力屬性，在特定的事項中具有專業的認定空間，但這並不代表公部門能夠不受原則的拘束，而運用無法順應法律續造需求的決策工具；最後，本文承認法治原則的內涵有調整的可能，但本文以尊重人的多元性作為法治原則調整的界線，並以人工智慧的運用已超出此邊界為理由，否認法治原則應隨人工智慧的使用而加以調整。換言之，在法治原則仍應具有規範性的地位下，若要使用人工智慧從事適用法律、作成決定的任務，在法制建構上尤其應該積極的回應無法進行法律續造的特色。

doi:10.6342/NTU202103631

最後，本文將於 4.3 中主張「受人為決定」與「人為介入」的管制方式，是公部門正當使用人工智慧的必要條件。相比於人工智慧，人類能夠在限制的條件下，掌握說的通的因果關係，透過反事實的假設以及逆推法，推論事物之間的合理關聯，並且透過文字加以形塑，這是人類相較於人工智慧的不同，也是人為決定與人為介入的管制措施下，以人類作為決策歸責對象的原因（4.3.1）。在說明本文所主張手段的正當性後，本文必須回答人為介入的要求下，人類如何有可能與人工智慧進行正當的互動（4.3.2）。本文主張建構以人工智慧設計、建置內涵為揭露對象的法制框架，透過揭露設計人工智慧的相關資訊，使公部門的決策者有能力分析，在個案之中是否採納或推翻人工智慧所提供的建議，避免使人淪為人工智慧下的奴隸。

doi:10.6342/NTU202103631

4.1　法律續造與法治原則間的關係

本節從法治原則出發，分析法律續造的操作方式，以及其在法治原則中的地位。[1]

4.1.1　法律續造的實踐

法律具有一般性（generality）的特色。[2]一般性指的是法律在制定上必須具有一定的普遍適用性，若非有特別的原因，否則不得以特定的對象，作為規範的對象。在一般性的要求下，法律者的制定者會以「種類」、「特徵」的方式，以形塑所欲規範的對象。假如，立法者希望能夠達成 φ 的目的，則透過一個規範 R，禁止具有同時具有 X、Y、Z 的特徵的行為。在這樣的立法中，φ 與手段「禁止 X、Y、Z 從事特定行為」之間具有促成可能性時，我們會說這個法律是一個具有理據正當的法律。

然而，法律是立法者在一時一地基於特定的知識脈絡下所創設的規範，必然會遺漏掉某些事物類型，造成應該管卻沒管、或不該管卻落入立法範圍之內的情形。在這種情形，若單純的依據法律的內容，對個案進行適用，則會發生有問題的結果。例如具有 K 特徵，但無 X、Y、Z 的對象，也同時會造成違反 φ 目的的結果，但法律並未對於此種情形加以管制，形成涵蓋不足（under-inclusive）的結果；抑或，管制對象雖然具有 X、Y、Z 的特徵，但不禁止其從事 R 規範下的行為，也不會發生 φ 的結果，而形成涵蓋過廣（over-inclusive）的情形。[3]這些情況皆意味者，R 的手段（透過管制 X、Y、Z），在某些情形，沒有辦法達成 φ 的目的，而在理據上不具有正當性。此時，法律適用者是否應該透過詮釋的行為，改變法律 R 的意涵，來促成 φ 的目的，即

[1] 在用語上「rule of law」在中文學界皆譯為「法治原則」，相較於此另一個字眼「Rechtsstaat」則為法治國原則。在淵源上，法治原則來自於英美法，著重的是普通法系下由法院透過正當法律程序承擔「人權保障」的任務；相較於此，法治國原則則是來自於德國法，以國會、立法機關制定的法律，來確保個人的權利與自由不受干預。參見黃舒芃（2009），〈法律保留原則在德國法秩序下的意涵與特徵〉，收於：黃舒芃編，《民主國家的憲法及其守護者》，頁14-15、註12，臺北：元照。由於本文後續論證上，尤其強調法院對於法律適用過程正當性的擔保，因此在意涵上指涉的對象即為「法治原則」，因此後續本文皆以法治原則作為說明的方式。

[2] See LON FULLER, THE MORALITY OF LAW 46-49 (1969)

[3] *See* FREDERICK SCHAUER, PLAYING BY THE RULES A PHILOSOPHICAL EXAMINATION OF RULE-BASED DECISION-MAKING IN LAW AND IN LIFE 31-34 (2002)

doi:10.6342/NTU202103631

涉及了法律內法律續造的問題。從此處的說明，可以得知法律續造尤其會發生在艱難案件（hard case），即單純依照文義在個案中適用，即便考量了所有的要素，但單純基於文義的規範適用將產生不妥適的結果。[4]

以上的情形涉及的是，文義解釋與目的解釋相互衝突時，應該如何解決的問題。在涵蓋不足的情形中，即是在文義解釋在個案的適用上有所不足，法律適用者是否應該透過以目的解釋的方式，增補法律並未規範的要件；而涵蓋過廣的情形中，則是文義在個案的適用上超越了立法者的目的，而法律適用者是否應該透過解釋，使案例事實例外的無法涵蓋於法規之中。如果要更仔細的討論這邊的問題，目的解釋中，除了立法者主觀的目的外，還包括客觀的目的，即法律秩序下所要求的特定目的。換言之，法律續造涉及的是文義解釋、主觀目的解釋、客觀目的解釋之間的衝突應該如何化解的爭議。

以本文於 3.3.1 中的說明可以得知，以上解釋要素之間並沒有嚴格要求的順序，僅有推定的順序。單從這三個解釋要素之間的衝突，第一個可以提出來的順序推定要求是：**文義解釋以及主觀目的解釋在個案的使用上，優先於客觀目的解釋**。[5]理由在於尊重立法者的權威地位。基於民主原則、權力分立以及法安定性的考量，立法者所制定的規範，透過基於目的的解釋或文義的解釋，拘束者法律適用者。然而，文義解釋相較於立法者的主觀目的具有公開性以及法安定的要求，所生的第二個順序推定要求是：**文義解釋優先於主觀目的解釋**。[6]

以上這兩個關於解釋要素之間的推定順序，由於僅是「推定」的順序，所以仍有推翻的空間。[7]依據法律論證理論的說明，要推翻基於規範文義的文義解釋必須負擔「論證」、「說理」義務，透過提出理由以正當化推翻推定的原因。亦即基於何種「理由」，法律適用者有必要違背法律文義，作成決定。

---

[4] *Id.*

[5] 參見王鵬翔（2004），〈目的性限縮之論證結構〉，《月旦民商法雜誌》，4期，頁27。

[6] 參見王鵬翔，前揭註5，頁27。

[7] 參見王鵬翔，前揭註5，頁28。

doi:10.6342/NTU202103631

隨著以上不同推定規則背後的不同法理，所需的論證強度以及正當性有所不同。以文義解釋與主觀目的解釋優先於客觀目的的解釋這個推定順序為例，文義解釋與主觀目的解釋之所以在順序上優於客觀目的解釋，最大的原因是「立法者具有權威地位」的理論基礎所支撐，而法律本身的內容是否正當僅在其次，這就是前文所說的「獨立於內容的理據」。[8]而法律適用者若要在個案之中採取客觀目的解釋，意味著支持特定解釋在內容上的正當性，必須強於這個支持權威地位的基礎。[9]

相較於此，「文義解釋優先於主觀目的解釋」這一個推定順序在推翻上就容易許多，理由在於不論是文義解釋或主觀目的解釋，都是基於「應尊重立法者」的法理所生的解釋方式。因此，只要基於主觀目的解釋的對於法規內容調整的正當性與平等原則無違，且不造成法安定性的破壞時，主觀目的解釋的優先性即可高於基於文義解釋。[10]

就以上的說明，我們可以得知透過法律續造在實踐上，並不會因為是一個個案的決定，而變成一個恣意的結果。[11]即便個案的法律適用者，在適用法律上遇到了一個艱難的案件，這也不代表法律續造的結果隨之發生。相反的，法律適用者必須提出的理由，除了規範內容上在正當性的質疑之外，還必須有能力跨越基於權力分立、民主原則所生的立法者權威。就此，我們可以說，一個合乎論證理性要求的法律續造，會隨著「權力性質的差異」而有不同的發生可能性。若要推翻立法者的權威，使客觀目的解釋優先文義解釋或主觀目的解釋，不僅於在權力分立上具有制衡的關係，且制度上也同時必須有一定的安排，例如法規合憲性的審查權的賦予。基於此，司法機關才有權限能夠從事推翻文義解釋與主觀目的解釋。相較於此，行政機關在權力分立上並

[8] 參見王鵬翔，前揭註5，頁28。

[9] 參見王鵬翔，前揭註5，頁29。

[10] 關於此處憲政法理學的討論請參見顏厥安（2002），〈規則、理性與法治〉，《臺大法學論叢》，31卷2期，頁1-58。

[11] *See* Reuben Binns, *Human Judgment in algorithmic loops: Individual justice and automated decision-making*, REGUL. GOV. 1, 3-4 (2021). *Contra* Frederick Schauer, *Is Defeasibility an Essential Property of Law*, *in* THE LOGIC OF LEGAL REQUIREMENTS: ESSAYS ON DEFEASIBILITY 85-88 (Jordi Ferrer Beltrán & Giovanni Battista Ratti eds., 2012); Juan Bayón Carlos, *Why is Legal Reasoning Defeasible?*, *in* PLURALISM AND LAW 340-343 (A. Soeteman ed. 2001).

doi:10.6342/NTU202103631

非制衡立法者的角色，也沒有制度上的措施，行政機關理論上僅能落實立法者的目的，推翻文義解釋而從事主觀目的的解釋。

4.1.2 法律續造與法治原則

「法治（rule of Law; *imperium legum*）」一詞，在概念上相對於人治，要求國家的各種權力必須以法律為依歸、受到法律所拘束。在此意義下，法律並非統治者的工具（rule by law），而是保護人民免於受到國家恣意壓迫的防護機制。雖然法治原則在意義上的高度歧異，不過在現代的發展上，法治原則在細部的發展上，可以分為更細緻的子原則，如形式的法治原則以及實質的法治原則。在形式法治原則上，法律的內容必須事前加以明確，人民可以可以理解以及預見特定法律適用的結果（明確性原則），不得對於已發生的事實的評價加以更動（禁止溯及既往原則）。除此之外，在實質上，法律作為國家權力落實的手段，必須有助於達成正當的目的，並且手段目的之間必須具有關聯性、不恣意，選擇達成目的最小侵害的手段，並且受影響的利益以及所達成的目的不得顯失均衡（比例原則）造成自由的侵害。

法律續造在法治原則中的地位，尤其視法治原則中的「法」究竟為何而定，除了實證法律之外，是否包括一定的正義的要求。換言之，在一個同時接受原則以及規則作為法律內涵的體系中，必然會承認法律續造的必要性。[12] 但這究竟是哪一種的法治觀呢？

法治原則的內涵除了強調法律在形式上必須符合一定要求的形式法治原則，與內容上要求一定正當性的實質法治原則外，法治原則另外一個比較少被提及的面向是強調「法律適用的過程（legal process）」，透過要求法律適用過程的正當，促成法治原則的落實。[13]這個面向的法治原則可稱為「程序的法治原則（procedural rule of law）」。[14]

---

[12] Karl Larenz（著），陳愛娥（譯）（2013），《法學方法論》，初版九刷，頁279，臺北：五南。

[13] *See* Richard H. Fallon, Jr., *"The Rule of Law" as a Concept in Constitutional Discourse*, 97 COLUM. L. REV. 1, 18-21 (1997).

[14] *See* Jeremy Waldron, *The Rule of Law and the Importance of Procedure*, *in* GETTING TO THE RULE OF LAW 5 (James E. Fleming ed. 2011).

doi:10.6342/NTU202103631

詳細的來說，憲法的法治原則既然是對於國家一切權力行使的拘束，可以合理的得出國家「適用法律」的行為也必須受到法治原則的拘束。最直接的要求，就是透過形式法治原則的落實，透過要求法律必須明確、事先規定，使人民可以預見法律的規範內容，並對於生活的行為加以安排。然而，從第三章對於法律適用的內涵的說明中，可以發現法律的適用本身並不是那麼的「固定」。以文字為基礎的法律，由於內涵無法確定，因此法律的適用過程中，法律適用者必須對於規範內容進行解釋，解釋的結果隨著個案的不同，在詮釋下規範的內容發生改變。此外，法律的論證具有可擊敗性，新的事證導致原先的法律主張被擊敗。規範的內容因為事過境遷，必須因為「法律原則」的介入，而有重新反省、甚至法律續造的可能。就此而言，法律適用的過程中充滿了不確定性以及開放性。此種規範結果的無法事前確定，伴隨著現代行政國家的出現，必須針對多元個案的個別適用法律以達到個案正義，法治原則在此時應如何發揮功能產生了疑義。[15]

如果說法治原則必須對於法律適用的過程也發生拘束，其產生拘束力的方式就是要求法律適用的過程必須符合「一定的準則以及方式」，而符合此種要求的法律適用過程，才是正當的法律適用方式。這也是以法律論證作為法律適用的方式，與一般論證的差異所在：法律的論證必須受到法律的拘束。[16]因此，針對同一命題的多數法律論證之中，應該以「何種標準」作為評價依據，才能最有效的達成普遍性的說服目的，尤其是法律論證理論中著重的焦點。但相關研究之中，不論是正面界定標準的相關嘗試[17]，或從法律論證的程序要求[18]出發建立法律論證的規則，最終面臨的共同困境在於僅能透過選擇立場或意識形態，在多數以「理性」作為論證基礎、具有相同說服力的相異論證中做出抉擇。[19]基於此困境，更進一步的一種觀點是，法治原則中的「法」本就包含法律可辯駁性的特質，因此法治原則與法律的可辯駁性本身並不互

[15] *See* NEIL MACCORMICK, RHETORIC AND THE RULE OF LAW: A THEORY OF LEGAL REASONING 12-16 (1 ed. 2005).

[16] ROBERT ALEXY, A THEORY OF LEGAL ARGUEMENTATION 217-218 (Ruth Adler & Neil MacCormick tans., Oxford University Press 1 ed. 1989) (1978).

[17] MACCORMICK, *supra* note 15, at 17-20.

[18] *Id*. at 21-23; ALEXY, *supra* note 16.

[19] MACCORMICK, *supra* note 15, 19-20.

doi:10.6342/NTU202103631

相衝突；在此前提中，法安定性或者法明確性的本質是建立於「可爭論」的性質之上，透過「訴訟」對於法的內涵進行澄清，達到法治原則。[20]

本文無意提出判斷標準以正面解決這個問題。不過，本文想指出的是，這些理論具有兩個共同的基礎。第一個基礎是，法律適用的結果必須依據個案的內容而定，並不是一個一致性、機械性的結果。法治原則視此種因為個案特殊性而從事的法律適用行為，為必然且必要。因此，在一般人為理性的法律適用導向的基礎外，無法透過其它限定特定的法律解釋結果的要求，以落實法治原則對於適用法律機關的權力拘束。

第二個基礎是個案中的法律論證，是否理性、是否符合論證上程序要求，或者以訴訟的過程作為法治原則的實踐等的不同理論，這些理論評價一個法律適用行為是否符合法治原則下理性標準的「標的」，是整個法律適用的過程，也就是法律論證所形塑的「理由」，透過分析個案中的理由是否理性、理由的形成過程是否符合論證上的一貫、在訴訟上分析理由的內涵是否可以支持，確保法治原則對於適用法律的過程加以拘束，並達到限制國家權力的目的。唯有透過法律適用者對於法律論證過程的描述，才能有效的評價法律適用的結果是否受到法律拘束。針對以上這兩個基礎，我們可以得知此種以「理由的給予（reason-giving）」為檢視法律適用是否正當的觀點，即是以法律適用過程為關注重點、著重個案中法律實踐的法治原則。[21]法治原則下的理由給予義務，相當於要求適用法律的機關將個案置於法律之中加以考量，達到透過法律限制國家權力拘束的法治目的。

相較於形式法治原則確保法律的形式內涵精確、實質法治原則確保法律在內容上的實質正義，著重法律適用過程的法治原則，強調的是法律適用過程的正當[22]，透過「程序正義」的方式落實法治原則的要求[23]；再者，此種法治原則是否被落實的基準，取決於個案中的法律適用對於一般人而言是否具有合理性（reasonableness）[24]；而在實踐上，則是透過個案的「理由提供」，

[20] *Id*. at 26-27.
[21] *See* Ashley Deeks, *Secret Reason-Giving*, 129 YALE L. J. 612, 627 (2020).
[22] *See* Mathilda Cohen, *The Rule of Law as the Rule of Reasons*, 96: 1 ARSP 1, 4-6 (2010).
[23] *See* Fallon, *supra* note 13, at 18.
[24] *See id*.

doi:10.6342/NTU202103631

將個案與法律加以連結，以確保個案的特殊性在法律的適用過程中加以考量[25]；最後，在制度上，尤其要求獨立司法制度的存在，透過事後的監督，擔保法律適用過程的正當以及正義[26]。

***

從以上的論證，本文從法律續造的方法以及許可性出發，釐清何種法律適用者在何種情形下能透過何種方式落實法律續造。此外，本文指出同時承認原則以及規則的法體系之中，必然隨著法律續造的要求。而法律續造的實踐也是注重法律適用程序的法治觀的內涵。換言之，本文以上的說明除了證成法律續造在憲法法治原則中的地位之外，也說明了法律續造在現實中實踐的可能與樣態。

[25] *See id*.

[26] *See* J. Harvie Wilkinson lll, *The Role of Reason in the Rule of Law*, 56 U. CHI. L. REV. 779, 784-787 (1989).

doi:10.6342/NTU202103631

4.2 法治原則下的自動化決定

在說明了法律續造在法治原則中的角色後，本節回到人工智慧的脈絡中，分析適用法律的主體，若由人類轉變為人工智慧，是否有不用進行法律續造的特殊原因。

部分贊成透過人工智慧作成個案決定者，其理由在於人工智慧，相較於人類決定，能更專業的考量個案情況、作成更正確的決定。[27]除了在現實中，大多數使用自動化決定的公務員完全不瞭解自動化決定運作的原理外[28]，人工智慧是否能作成更專業的決定也有疑問。人工智慧在開始運用之後，透過資料所訓練出的模型就已經固定。固定的模型考量的因素僅限於訓練資料中具有的資料類型，超出訓練資料的「真實世界」即因此沒有辦法被納入考量的可能，因此人工智慧即有很大的可能漏未考慮應考慮的因素[29]。此外，資料之間的關聯性，可能正確的代表社會生活之間的事物關聯，但也可能根本的倒果為因的錯誤認定事物關聯。若運用後者從事適用法律作成國家決定的任務，反而會造成錯誤的適用法律，而非更加專業的法律適用。

另外一種可能的支持見解是，若僅是一個重複發生的案件，人工智慧仍有能力能夠作成正確的決定。本文不同意這樣的見解。首先，決定的正確性並不是人工智慧說了算，而是要由第三者加以驗證。因此使用人工智慧的行政機關若要說明其權衡正確，必須說明自動化決定究竟考量了哪些因素，以及如何考量。然而行政機關通常不是人工智慧的建置者，因此自己並不知道人工智慧背後究竟考量了哪些訓練資料。此外，即便知道人工智慧的系統考量了哪些要素，模型本身也沒有辦法說明個案中的每個要素的重要性。更進一步來說，人工智慧根本沒有權衡的能力。如第三章所指出的，演算法所生成的模型除了沒有人能夠理解之外，模型本身的原理僅是利用了巨量資料所

[27] *See*, *e.g.*, Felix Mormann, *Beyond Algorithms: Toward a Normative Theory of Automated Regulation*, 62 B. C. L. REV. 1, 4 (2021).

[28] *See* Ryan Calo & Danielle Keats Citron, *The Automated Administrative State: A Crisis of Legitimacy*, 70 EMORY L. REV. 797, 818-831 (2021).

[29] *See* Marion Oswald, *Algorithm-assisted decision-making in the public sector: framing the issues using administrative law rules governing discretionary power*, 376: 20170359 PHIL. TRANS. R. SOC. A 1, 13-14 (2018).

doi:10.6342/NTU202103631

呈現的「歸納法」運作模型，沒有辦法區分個案的「特殊差異」並作成權衡的決定[30]。換言之，人工智慧既然沒有權衡的能力，也不可能正確的作出權衡。

不過，以上針對贊成人工智慧在公部門使用主張的回應，僅能說明人工智慧在表現上，與理論上的法治原則下的法律續造要求有所不符，但在個別決定的「多數結果表現」上，人工智慧並沒有遜於現行適用法律的人類。換言之，在精確性上的要求，即便人工智慧沒有納入新的資料、無法針對新事實進行判斷、無法進行權衡，也只不過是跟目前多數的法律適用者相當，雖然比上不足，但仍舊比下有餘，某程度仍舊能夠取代人類從事法律適用的任務。在這樣的觀點下，從技術與法律之間的互動加以切入，更核心的質疑是，法治原則也僅是達成正義或者人權保障的一種手段，而非僵固無法改變的鐵則。在面對社會、經濟與文化所促成的自動化決定，法治原則的本身毋寧應該隨著科學與技術的現實加以微調，從「法治原則 1.0」提升至「法治原則 2.0」。[31]

本文同意法治原則有改變的可能，甚至應該與時俱進，但法治原則所欲達成的「避免國家恣意」與「保障人民權利」的目的，並不會因此有所改變。而這些以「人類」為保護中心的目的要求，也為人工智慧的使用劃下不可撼動的界線。人類的特殊性在於每個人之間有差異，這些差異造就了多元的人類，並且形塑了多元的社會。國家權力的行使中，必須注意這些差異，將這些差異納入考量。傳統意義下的法治原則，以「文字」作為法律內涵的技術載體，因為文字所生的規範內涵不確定性，因此必須透過法律專業人士對於法律文字進行解釋，將個案情況的差異規範的解釋過程之中，個人的差異因此被國家加以考量以及重視。[32]換言之，法律因為文字作為載體所生的「不確定性」，能夠被法律適用者以個案理由、以理性一般人的角度下「說的通」的因果關係[33]，將個案的脈絡，宏觀性的將不同的事物納入法律適用的過程，

[30] 請參見本文第二章3.4.1以下的論述。

[31] *See* JULIE E. COHEN, BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTIONS OF INFORMATION CAPITALISM 4, 247-248 (1 ed. 2019).

[32] *See* MIREILLE HILDEBRANDT, SMART TECHNOLOGIES AND THE END(S) OF LAW 181-183 (1 ed. 2015).

[33] *See* KENNETH CUKIER, VIKTOR MAYER-SCHÖNBERGER & FRANCIS DE VERICOURT, FRAMERS: HUMAN ADVANTAGE IN AN AGE OF TECHNOLOGY AND TURMOIL 60-63 (1 ed. 2021); Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORDHAM L. REV. 1085, 1096-1098 (2018).

doi:10.6342/NTU202103631

透過理性的理由說明，進行法律解釋以及適用，對於個案的差異形成外界能夠理解的個案決定。[34]在這種方式下，現行法律體系透過法律續造的方式，能夠面對個體的差異，作出具有合理差異的決定。換言之，我們可以說，任何程度的法律續造的要求，以及保留法律續造的可能性，在規範上都是必要的。

當然，「文字」的技術不應該決定法治原則的內涵，但如果要從「文字的法治原則」進入到「資料的法治原則」，必須要考量的因素是個人與個人之間的差異以及多元性，是否能就能夠被法律適用者正當的考量。遺憾的是，人工智慧沒有辦法針對個案的差異而適用法律。人工智慧的運作，以資料之間的關聯填充社會事實的種種不確定，以達到客觀以及穩定的運作，但在人工智慧消除不確定的過程的時候，也將每一個受決定者的差異加以去除。[35]自動化決定利用數據在過往多數案件中的角色，對於現今的單一個案形成決定。表彰特定破碎事實的資料，具有去脈絡化的特色[36]，若以資料之間的關聯運用於決定的過程，相當於並未將個案的脈絡、事物與事物之間的關係以及特殊性納入法律解釋、個案涵攝之間的考量，僅是因為特定特徵與結果的資料之間的關聯性，就作為適用法律結果的依據。這種忽略人類個體差異的國家權力行使，與其說是未來法治原則的樣貌，不如說是法治原則自始要避免以及限制的對象。

總結以上的說明，法律續造之所以在公部門具有必要性，從規範上來說，是基於「原則」作為某種正義載體的實現，從理論上來說，則是看待每個人的差異而作出「個案」決定。因此，在法治原則的要求下，人工智慧的「直接」使用，尤其因為其不具有法律續造的可能性，在規範上就法治原則有所違背。

---

[34] *See* Kiel Brennan-Marquez, *"Plausible Cause": Explanatory Standards in the Age of Powerful Machines*, 70: 4 VAND. L. REV. 1249 (2017).

[35] *See* SUN-HA HONG, TECHNOLOGIES OF SPECULATION: THE LIMITS OF KNOWLEDGE IN A DATA-DRIVEN SOCIETY 16-29 (1 ed. 2020).

[36] 參見Viktor Mayer-Schönberger（著），林俊宏（譯）（2015），《大數據：隱私篇》，1，頁106，臺北：天下。

doi:10.6342/NTU202103631

4.3　人爲介入爲核心的人工智慧管制框架

從 4.2 的說明，本文分別指出國家若使用人工智慧直接進行適用法律、作成決定的任務時，將與法治原則的要求有所相違背。那公部門在何種條件下，能夠使用人工智慧來從事公部門的任務，而不會違反法治原則的要求呢？

學說上對於「演算法可課責性（algorithmic accountability）」的討論，主要在於透過一定法律制度的建立，一方面要求演算法工具的使用者必須說明，為何使用演算法工具具有正當性，另一方面則是確保受決定者有能力改變產生不利影響的演算法決定，或不受到演算法決定的不利影響。[37]若將法治原則與對於演算法可課責性加以連結，可以清楚的得知，公部門之中演算法的課責性的確保相當於：如何使公部門使用演算法工具（人工智慧）符合法治原則。

一種簡潔並且符合憲法要求的自動化決定管制模式的方法就是完全禁止自動化決定於公部門之中的使用。然而，人工智慧在公部門之中，仍舊有效率以及一致決定上的誘因。[38]本文並不認為這樣的誘因是不正當的，只是若要使用人工智慧，應該以符合法治原則的方式加以使用。

本文認為透過「人為介入」的管制方式，一方面能夠避免人工智慧的使用違反憲法法治原則的要求，除此之外還可以使國家享有人工智慧所帶來的利益。以下，本文於 4.3.1 說明為何「人為介入」能夠使人工智慧的使用符合憲法要求，並於 4.3.2 主張建立以人工智慧的模型設計為揭露對象的管制模式，以有效的落實「人為介入」的規範性要求。

4.3.1　人爲介入的正當性

若以本文於第一章所提出的歐盟一般資料保護規則中的「人為介入」作為討論中心，「人為介入」在行使上，原則上能夠使個案的當事人避免受到自動化決定的效果加以影響，取而代之的是人為的決定；例外的情形中，雖

[37] *See* Talia B. Gillis & Joshua Simons, *Explanation < Justification: GDPR and the Perils of Privacy*, 2 J.L. & INNOVATION 71, 77-80 (2019).

[38] 例如不同的人類經常在同一種案件中，作成不同的決定，學者稱此為「雜訊（noise）」。*See generally* DANIEL KAHNEMAN, OLIVIER SIBONY, CASS R. SUNSTEIN, NOISE (2021)

doi:10.6342/NTU202103631

然仍舊會受到自動化決定的影響，但仍保持著人為介入的機會。透過保留最終「人類決定」的方式，人工智慧無法個案適用法律並突破法律框架的問題，並不會出現於最終的個案決定中。取而代之的是人類適用法律的結果。人類必須透過對個案的觀察，適用法律，在考量個案是否有特殊性後，透過理由的呈現，作出最終的決定。

不可避免的是，人類會做錯決定。但人類提供的理由，能夠作為後續檢視錯誤的機制。此外，理由的內容，使個案的受決定的對象能夠明確的瞭解法律之所以對於「特定個案」產生「特定法律效果」的「特定原因」。人類專有的「權衡」的能力，尤其能夠在個案之中實現，透過判斷個案之中是否有相較於過往案例的特殊的情形，並且分析此特殊情形是否應透過「原則」的適用而達到突破法律框架的結果。因此，人為決定的優點以及必要性，就在於透過理由的呈現，以及對於個案進行正當的評價，以符合法治原則的要求。

如果要正面的來說明「人為介入」能夠使適用法律作成決定符合法治原則要求的原因，主要的原因在於人類能夠真正的理解世界，並且以人類的思考方式以及觀點，說明事物之間的關係。就適用法律作成決定的任務而言，必須包括：解釋法律、個案適用的步驟。針對規範進行解釋時，規範的意涵並非取決於獨立字元的特定文義，而是取決於特定規範的脈絡加以決定。僅有具有世界觀、理解事物之間關係的人類，才能夠過濾正確的資訊，以釐清特定脈絡下的文義。[39]另外，在調查個案事實是否符合現實的規範情形中，必須就所獲得的有限證據進行拼湊以及評價，而建立出對於事實狀態的合理假設，以形成法律事實。從單一的證據到整體事實的宏觀圖像，仰賴的也是人類對於世界的整體認識，單一的證據（數據）無法直接的決定個案的法律評價。[40]在面臨到新的個案的事實時，人類能夠透過掌握對於事物之間的抽象因

[39] *See* ERIK J. LARSON, THE MYTH OF ARTIFICIAL INTELLIGENCE: WHY COMPUTERS CAN'T THINK THE WAY WE DO 191-234 (1 ed. 2021). Larson以「逆推法（abduction）」說明此種只有人類能夠從事、但人工智慧無法從事的推論方式。所謂的逆推法指的是就所觀察到的特定現象，給予其個案性的合理理由的解釋或者假設。因此，逆推法並非一般性的，而是個案性的。此外，該種理由或者假設的生成仰賴的是聯想、常識，亦即有道理的猜測。但也因為是假設或者猜測，逆推法具有可擊敗性，事件的脈絡會影響假設的理由是否成立。逆推法仰賴的能力是體系性的觀察以及對於事物的真實理解，而這兩個能力的缺乏也直接造成人工智慧無法進行逆推法。

[40] *See* Pek van Andel & Daniele Bourcier, *Serendipity and Abduction in Proofs, Presumptions and Emerging Laws*, *in*

doi:10.6342/NTU202103631

果關係，並且進行反事實的思考，釐清新個案在現行法律框架上的定位[41]。因此，唯有真正理解世界的人類，才能整體的考量個案之中的情形，作成個案的決定。

在釐清「人為介入」在理論上的正當性後，本文要更進一步的說明人為介入在個案之中應該如何設計。人為介入與截然禁止人工智慧的使用，兩種管制方式的差異在於，前者仍舊保持了人工智慧在決定過程中的使用可能性。本文在第二章對於人為介入的說明中，指出人為介入的內涵中至少包括了兩種選項，「輔助人為決定型」與「事後介入保留型」。在這兩種意義對於人工智慧的使用分別具有不同程度的限制，前者相當於禁止人工智慧對於個人發生影響，而後者則是容許自動化決定對於個人直接產生法律效果。在現今國際的浪潮廣泛推廣人工智慧的情形中，尤其可能會廣泛的採取後者，追求快速的作成正確決定。

但是，雖然採取人類事後保留的見解能夠達到「人類決定」的目的，但在現實中，我們是否能夠期待當事人「主動的」行使人為介入的「權利」，而避免遭到自動化決定的侵害？從個人資料保護法「同意原則」在網路時代的實踐會發現，賦予當事人自主的同意權，在網路習慣以及商業模式的情況中，最終反而會造成「同意啦，哪次不同意」的無意義結果。[42]我們可以合理推測，自動化決定的情況中，當事人自我放棄權利的情況可能再次出現。另外，人類事後保留的機制，相當於當事人在行使人為決定權利之前，必然會先受到自動化決定的效果所及。因此，此處所發生的另外一個問題就是，是否應該基於快速以及節省人力的考量，將此「程序成本」轉交給受決定者加以承擔？[43]

THE DYNAMICS OF JUDICIAL PROOF: COMPUTATION, LOGIC, AND COMMON SENSE 278-81 (Marilyn MacCrimmon & Peter Tillers eds., 2002).

[41] *See* CUKIER, MAYER-SCHÖNBERGER & VERICOURT, *supra* note 33, at 56-59, 86-87.

[42] *See* Lilian Edwards & Michael Veale, *Slave to the Algorithm: Why a Right to an Explanation Is Probably Not the Remedy You Are Looking for*, 16 DUKE L. & TECH. REV. 18, 33-35 (2017); HELENA U. VRABEC, DATA SUBJECT RIGHT UNDER THE GDPR 213 (1 ed. 2021).

[43] *See* Ari Ezra Waldman, *Power, Process, and Automated Decision-Making*, 88 FORDHAM L. REV. 613, 613-632 (2019).

doi:10.6342/NTU202103631

考量到「事後保留人類介入」的規範模式在現實中的運作，可能完全架空人類決定的規範目的，甚至間接的導致人為介入要求下的人工智慧仍舊與法治原則有所違背的結果，本文認為在模式上應該更細緻的區分事務領域對於人為介入的內涵加以設計。如同本文前面所指出的，人為介入對於法治原則的傷害主要來自「無法對於新個案適用法律」以及「法律續造」。在此前提上，本文前述關於法律續造所發生的事務領域，以及特定領域的規範可能比較具有變動性、而需要個案決定時，即應在此處作為「人為介入」具體實踐的考量。

首先，著重法律適用程序的法治原則特別要求，司法機關對於個案法律適用的確保，並且必須落實法律的續造。並且司法機關的法律續造，除了基於立法者主觀的目的續造類型之外，尚及於基於客觀的目的進行法律續造。因此在法院的情形中，若使用人工智慧，僅能使用「輔助型」的人工智慧，必須由人類作成每一個最後的決定，而不能由自動化決定取代人類適用法律作成決定的任務。

再來是行政機關的情形，行政機關在國家中從事著不同的任務，而從法治原則對於行政的要求下，尤其必須考量所涉及事務領域、以及所涉及的法律要件性質來區分如何運用「人為介入」。

若特定法律的結構多為「確定的法律概念」所組成，並且皆為「羈束」的法律效果時，由於法律解釋的空間已經被立法者加以限制，因此對於行政機關而言，此時已經不太具有透過法律續造推翻規範內容的正當性。因此，在此時能夠採取「事後保留人為介入」，原則上容許使用自動化決定，而在事後透過人為的介入擔保個案決定的正當性。例如：「時速超過每小時 60 公里，罰 1600 元」，在這樣的規範中，由於要件皆由確定法律概念加以組成，並且立法者限定了確定的法律效果，因此能夠使用自動化決定。

此外，在大量行政的案件之中，如稅務案件，由於此種案件中尤其具有效率的需求，並且多透過明確、制式的法律文義形塑規範，個案並不太會在行政階段中有超出法律解釋而續造的需求，因此得採取「事後保留人為介入」的方式來運用人工智慧。

doi:10.6342/NTU202103631

但在此範圍之外的行政決定，國家若使用人工智慧來作成決定即應該採取「輔助型」的使用。理由在於，國家行政具有執行力、強制力的特徵，在國家行政行使的狀況之中，不應使人民承擔「事後保留人為介入」下對於自動化決定不服而去挑戰的程序成本。因此，不得以自動化決定的方式作成直接的決定，僅得透過人工智慧輔助人類為主的行政機關作成最終的決定。

茲將國家各種決定與中使用自動化決定的容許條件條列如下：

- 司法決定（例如：判決）：輔助人類決定型
- 行政決定
  - 確定法律要件、羈束效果：人類介入保留型
  - 大量行政事件（例如：稅務事件）：人類介入保留型
  - 其它行政事件：輔助人類決定型

綜合以上的說明，本文首先說明為何「人類介入」在實踐上為何能夠符合法治原則的要求，並且依據法治原則的內涵，以及人為介入在實踐中的考量，證成人為介入在實踐上更細部的規範要求。

4.3.2　以人爲介入爲核心的人工智慧管制框架

在證成人為介入在國家各種決定中應如何落實的規範性要求後，本文要進一步的說明人為介入應該如何落實，何種法制框架的建構協助人類判斷是否接受或拒絕人工智慧的建議，以符合課責的要求。

4.3.2.1　透明作爲人機互動挑戰的解方

「人為介入」的管制模式雖然保持著人工智慧在公部門中使用的可能性，但仍舊要求必須由人類作成最終的決定，將決定所生的問題溯源於人類的決策者，決策者必須說明為何作成特定的決定理由，達到課責的目的。因此，即便能夠使用人工智慧，但最終必須負責的對象是「人類」。

doi:10.6342/NTU202103631

但在文獻上不乏指出，人機互動的過程中，人類可能過度的相信電腦，而不相信人類的決定結果[44]；此外，在本文第三章對於人工智慧的特性的說明中，人工智慧的運作過程具有無法理解的特性。在這些特性的影響下，人類如何有能力決定是否接納人工智慧所提供的結果，或在人類介入保留的情形中，他人對於自動化決定加以挑戰時，人類應如何審查自動化決定的正當性，是「人為介入」在落實上必須重視的問題。針對這個問題，本文認為人類不應該對於現階段技術下的人工智慧抱持著「幻想」，機器並不是永遠都能作出比人類更正確的決定。在機器並不擅長的領域中，就會作成錯誤的決定，例如在資料不足、事實變化具有多樣性、特定的政策目標難以程式化等領域，都可以說是人工智慧在表現上可能差於人類決定的領域。因此，何時機器會出錯，尤其仰賴人類對於特定人工智慧內涵的掌握。

由於人工智慧輔助人類作成決定的正當性，取決於使用者對於特定人工智慧內涵的掌握，在規範上即應該以落實「透明（transparency）」的方式，透過揭露使決策者知悉人工智慧的內涵，以正確評估是否使用。在人工智慧治理的討論之中，「透明」手段是經常被討論的一種管制手段：透過透明，打開人工智慧的黑箱，避免自動化決定所生的法律問題。[45]透明指的是透過提供行為的相關資訊，供第三人能夠檢視行為的正當性，進一步達成監督的目的。[46]就此，我們可以得知透明制度若要有效，並不能僅有單純的揭露，而必須搭配後續的究責機制。[47]然而，透明所欲促成的目的有所不同，而不同的目的會要求揭露不同的資訊。因此，在利用人工智慧作成決定或者輔助決定作

[44] 學說上稱此為「automation bias」。*See* Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1271 (2007).

[45] *See, e.g.,* Tal Z. Zarsky, *Transparent Predictions*, 2013 U. ILL. L. REV. 1503 (2013); Margot E. Kaminski, *Understanding Transparency in Algorithmic Accountability*, *in* THE CAMBRIDGE HANDBOOK OF THE LAW OF ALGORITHMS (Woodrow Barfield ed. 2020); Cary Coglianese & David Lehr, *Transparency and Algorithmic Governance*, 71 ADMIN. L. REV. 1 (2019); FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION (1 ed. 2016).

[46] Nicholas Diakopoulos, *Transparency*, *in* THE OXFORD HANDBOOK OF ETHICS OF AI 198 (Markus D. Dubber, et al. eds., 2020).

[47] *See* Mike Ananny & Kate Crawford, See*ing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability*, 20 NEW MEDIA & SOCIETY 973, 974 (2018); Hannah Bloch-Wehba, Transparency's AI Problem (Knight First Amendment Institute and Law and Political Economy Project's Data & Democracy Essay Series, 2021, Texas A&M University School of Law Legal Studies Research Paper No. 21-13), available at: ogy.de/cuq1.

doi:10.6342/NTU202103631

成時，究竟應該揭露何種資訊方能有效的符合人工智慧的課責要求，即生疑問。

在人工智慧與透明的討論脈絡中，以「具說明能力的人工智慧（explainable Artificial Intelligence, xAI）」為中心的討論，尤其著眼於何種資訊的揭露有助於提升人工智慧本身的可信任性。[48]「具說明能力的人工智慧」指的是人工智慧在做成決定時，能夠同時透過一定的方式，協助人工智慧的使用者理解結果產生的原因。具說明能力的人工智慧是一個來自於資訊科學界的概念，而非因為法律的要求所出現的技術類型。對於資訊科學界而言，人工智慧的「可理解型」與「效能」處於一個互相衝突（trade-off）的關係，可理解的人工智慧預測效能並不是非常顯著，但不可理解的人工智慧，其預測效能即非常顯著。[49]為了增加效能高的人工智慧的應用與普及程度，資訊科學界以「具說明能力的人工智慧」作為研究方向，透過一定的資訊技述措施提升人工智慧的可理解性，以提升人工智慧的信任。就此概念在適用法律的任務上應如何評價，尤其應著重具說明能力的人工智慧的內涵究竟為何。

說明能力人工智慧的發展目前主要有兩種路徑，一種是解構模式（decompositional approach），另一種則是外部模式（exogenous approach）。[50]解構模式透過理解人工智慧的內涵，達到說明的目的。例如，揭露所使用的演算法或者訓練資料。不過這種方式在實務上基於營業秘密以及無法理解而不常被使用。另外一種解構模式，則是在想要理解的機器學習（A）下，再外掛一個較具有理解可能性的機器學習（B），以 A 的來源資料以及決定結果作為 B 的訓練資料來重建 A 的可理解性。換句話說，即從一個可以被理解的模型，詮釋另一個無法被理解的模型。這種類型的具說明能力的人工智慧的類型稱為「中介模型（surrogate model）」[51]。

---

[48] *See* Ashley Deeks, *The Judicial Demand for Explainable Artificial Intelligence* 119 COLUM. L. REV. 1829, 1833 (2019).

[49] *See* Pantelis Linardatos, Vasilis Papastefanopoulos & Sotiris Kotsiantis, *Explainable AI: A Review of Machine Learning Interpretability Methods*, 23(1): 18 ENTROPY 1, 1-2 (2021).

[50] *See* Deeks, *supra* note 48, at 1835.

[51] *Id*. at 1837-1838.

doi:10.6342/NTU202103631

相較於解構模式深入人工智慧本身的內涵，外部模式並未實際著眼於演算法決定的內部過程的說明，而是著重於透過外部資訊而後設的說明決定可能的運作方式。外部模式的說明內容可能是著眼於模型的資訊（model-centric）或者是著重於主體的資訊（subject-centric），前者又稱整體的說明（global explanation），後者稱為個別的說明（local explanation）。[52]模型本身的資訊包括：自動化決定設計者的目的、所使用的演算法類型、訓練資料的說明、如何測試訓練模型等資訊；而主體的資訊則諸如：哪一個資料內容的變動將導致個案決定的改變、相同條件的其他決定者的決定結果為何、獲得相同決定的其它個案的情形等情形。目前有特別討論的反事實的說明（counterfactual explanation），即著眼於哪一個資料內容的變動將導致個案決定的改變，例如當事人是月薪 32000 元而獲得自動化決定的不利解決，而請求說明其他條件相同，但申請人月薪是 48000 元的決定結果。[53]

在釐清技術透明手段的內涵後，必須進一步檢視的是這些資訊科學界所提出、用以提升信任的手段，是否有助於達成法治原則下的課責目的。在法治原則所推導出的人為介入，指的是決定者必須要理解人工智慧系統之所以產生特定決定的現實原因，透過評估這個現實原因是否與法律規範的內涵相符，進而決定是否使用、採納或者推翻人工智慧的決定。因此，相關手段至少必須提供使人類能夠「理解」並且分析決定的過程是否「與法律相符」的內涵。

若不考慮營業秘密的問題，解構模式中關於揭露人工智慧內涵的方式僅說明了個別決定會考量的要素（資料）以及這些要素形成運作規則的方式（演算法）。除了資料過於龐大的原因之外，單純這種技術資料的揭露，沒有辦法說明支配人工智慧運作的規則，也沒有辦法使外界理解每一種資料對於最終的決定有何種影響。[54]換言之，單純的提供這些技術性的資料，沒有辦法促使決定者更理解自動化決定的模型本身，也沒有辦法作成是否採納自動化決定的實質判斷。而中介模型的方式，也無法符合法治原則的要求。中介

[52] Edwards & Veale, *supra* note 42, 55-59.

[53] Sandra Wachter, Brent Mittelstadt & Chris Russell, *Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR*, 31 HARV. J. L. & TECH. 841 (2017).

[54] *See* Joshua A. Kroll, et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 657-660 (2016).

doi:10.6342/NTU202103631

模式的問題在於創造了信任的假象，理由在於，若透過無法理解的 A 模型做成決定，即便用比較能理解的 B 模型說明 A 模型的本質，也不能代表 B 模型就是 A 模型。[55]換言之，A 模型的無法理解性仍然不因 B 模型而終局的解決，因為 A 模型仍舊可能做出 B 模型不能解釋的決定結果。

而外部模式中，著重個別主體的個案說明提供關於「其它種情形中」的決定結果。本文認為這種資訊，並沒有辦法協助決定者判斷是否應該採納該系統的決定。對於法律適用而言，重要的是過程而並非結果，因此單純從兩個個案之中，特定因素的差異，即得出不同結果的情形，反而將法律的適用淪為機械性的操作過程。[56]在前述反事實說明的案例之中，月薪資不同而獲得不同決定結果的過程之中，即便我們知道是「薪資」的差異在最終導致了不同的結果，但其僅提供特定「個案」之中的說明功能。在這樣有限的資訊下，使用人工智慧的決定者，無法全盤的判斷整個系統的運作方式，以及是否適合適用於特定的決定領域。

從以上的說明，我們可以發現相關著重揭露人工智慧的「技術內涵」的透明手段，仍舊沒有辦法有效的打開人工智慧的黑箱、使外人理解支配特定人工智慧系統的運作過程。但本文認為：無法理解人工智慧的運作過程並不妨礙決定的使用者評估自動化系統的本身。在人為介入的規範性要求下，由於決策者並未將「權限」移轉至機器，因此課責的對象仍舊是決策者。[57]換言之，在決策者使用人工智慧時，法治原則對於決策者的要求即轉變為：**為何決策者在個案中使用人工智慧的決定方式是正當的**。決策者必須正面的說明為何在個案中透過人工智慧輔助所作成的決定，相較於樸素的人為決定更加正當。在這樣的課責要求下，決策者必須正面的審視特定人工智慧具有何種特質，該特質對於個案的決定會產生何種影響，並決定在個案之中是否適用特定的自動化決定。

---

[55] *See* Jake Goldenfein, *Algorithmic Transparency and Decision-Making Accountability: Thought for Buying Machine Learning Algorithms*, *in* CLOSER TO THE MACHINE: TECHNICAL, SOCIAL, AND LEGAL ASPECTS OF AI 58 (Sven Bluemmel ed. 2019).

[56] *See id*.

[57] *See* Jennifer Cobbe, *Administrative Law and the Machines of Government: Judicial Review of Automated Public Sector Decision-making*, 39 LEGAL STUDIES 636, 645-646 (2019).

doi:10.6342/NTU202103631

如同本文於第三章 3.1 中所指出的，人工智慧的本質是由資料、演算法以及各種人為的判斷所作成，在人工智慧仍不具備自我意識的情況下，仍舊是依據人類的指令加以行事。因此，前述的三個「素材」直接的影響了決定的結果。[58]在這樣的課責要求下，本文認為前述「**外部模式」下著重模型建置階段的相關說明措施，反而是符合法治原則要求下的管制手段**。透過提供關於自動化決定系統形成階段的各種資訊，供使用者加以分辨是否能夠有助於特定任務的達成，進一步決定是否採納、推翻人工智慧所建議的決定，有效的達成人為介入下人類與機器之間的良善互動。

人工智慧建置階段的相關資訊不必然是涉及技術本質的資訊，而是特定的人工智慧系統在不同階段之中，為了達成「何種目的」做了「何種設計上的選擇」，決策者透過分析這些「設計上的選擇」判斷其在規範上是否能加以容許。這些資訊在內涵上與技術本質資訊的差異在於，前者在人類社會生活之中具有意義，使用者具有整體評估特定人工智慧系統使用的容許性。因此，提供這些資訊的目的就在於確保使用人工智慧的決定作成者，能夠正確的視任務需求為何，選擇符合任務內涵的工具。換言之，這些資訊目的在確實的將法治原則的深入人工智慧之中，並不因為使用人工智慧就豁免於法治原則的拘束。[59]使用者針對資訊作出判斷，釐清個案中的任務是否能夠使用特定的人工智慧，而後續的監督者則審查自動化決定的使用者依據這些資訊所作成的「判斷」是否具有正當性。透過這種透明手段的落實，則能夠確保自動化決定的使用仍舊能夠達成課責的目的。

4.3.2.2　以設計內涵透明爲目的的法制框架建構

從上述的說明，我們可以得知針對模型形成的內涵所為的透明措施，是有效落實法治原則下人為介入的必要手段。本部分嘗試進一步的說明在人工智慧的建置的不同階段之中應該揭露何種涉及評估人工智慧內涵的重要資訊。

依據本文於第三章中 3.1 的說明，人工智慧的建置可以大別為「**定義目標→蒐集並且處理資料→透過演算法得出模型→調校→應用**」等的階段，這些

[58] *See* Deirdre K. Mulligan & Kenneth A. Bamberger, *Procurement As Policy: Administrative Process for Machine Learning*, 34 BERKELEY TECH. L. J. 781, 794-808 (2019).

[59] *See* Kaminski, *supra* note 45, at 127-129.

doi:10.6342/NTU202103631

階段的進行，都仰賴設計者作成不同的判斷，以形塑人工智慧的本質。以下分別說明各階段應揭露的資訊。

在**定義目標**的階段之中，建置者應該揭露特定的人工智慧系統所欲達成的目標為何，以及透過何種指標的追尋達成政策目標的需求。[60]在政策的面向中，許多政策的目的無法透過單一事物項目的數值加以明確的呈現，而是多數事物項目綜合出的結果，因此必須判斷特定人工智慧的目標與政策目的的吻合程度，若特定領域之中對於人民權利的影響愈高，系統的目標與政策的吻合程度即應愈高。

在**蒐集以及處理資料**的階段中，建置者尤其應該揭露所蒐集資料的來源以及蒐集方法、對於資料標籤的標準、所使用的變項/特徵以及其它對於資料處理的相關標準。由於資料直接的影響了決定的結果，所以對於資料的內涵尤其必須明確的說明。首先，揭露所蒐集資料的來源以及蒐集方法，目的在於決策者判斷以此資料為基礎所建置的人工智慧究竟能夠運用至何種個案之中以及判斷該人工智慧的偏差程度。[61]理論上來說，僅有以日常的情形為基礎、進行隨機抽樣所蒐集的資料，並以此資料作為人工智慧的基礎才足以反應現實。但大多數資料的蒐集方式，並非以隨機抽樣的方式加以蒐集，因此資料的內涵會與現實的情形產生偏差。除了必須避免特定錯誤偏差所導致的歧視結果外，尤其必須用以在使用的階段中判斷「個案」是否屬於涵蓋於所蒐集的「資料」範圍中。若特定個案的情形，並未如同所蒐集資料的蒐集於資料庫之中，以此資料庫所訓練的人工智慧即可能無法應對此種個案作成正確的決定。

再者，在資料處理的階段之中，另外一個必須揭露的事項是特定資料庫中所使用的資料變項，以及變項的內涵。人工智慧透過觀察變項與結果之間的關聯性作成決定，因此資料庫中有或沒有特定變項，都會影響決定的結果。揭露所使用變項主要的目的即在於，確保人工智慧考量了法律規定應考量的變項，不考量法律所禁止考量的變項。基於反歧視的要求，若資料庫中的變

[60] *See* Jennifer Cobbe, Michelle Seng Ah Lee & Jatinder Singh, *Reviewable Automated Decision-Making: A Framework for Accountable Algorithmic Systems*, PROCEEDINGS OF THE 2021 ACM CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY (FACCT '21) 598, 603-604 (2021).

[61] *See* Oswald, *supra* note 29, at 17.

doi:10.6342/NTU202103631

項涉及法律所禁止的特徵時，則有理由不使用該人工智慧的系統[62]；相反的，若法律要求應該考量特定因素，則在系統之中尤其必須考量相關的特徵，並作成正確的決定。[63]除了變項本身的揭露之外，變項下的內涵也要揭露。舉例而言，若以性別作為變項，性別單純以「男/女」的二元區分與「男/女/其它」的多元區分即對於結果產生不同的影響。[64]變項下的內涵是否足以含括社會的多元性，影響的是特定的人工智慧系統如何細緻的對待不同的差異以及差異與差異之間的關聯。若特定的領域之中，要求更細緻的區分時，特定的人工智慧系統則有可能不敷現實的使用。

在資料處理的階段之中，包括其它諸多細瑣的人為判斷，像是針對所蒐集的個別資料的標籤行為、區分訓練資料與測試資料、排除過度異常的資料等清理資料的行為…等其它的處理資料的行為，製造者應揭露這些行為的標準以及執行方式。從整體性的觀點來觀察，這些行為個別雖然都是微不足道的人為判斷，但若標準不一致，或一致的產生系統性偏誤時，相關人為判斷都會影響最終結果。因此，相關人為判斷的準則以及方式也有揭露的必要。

在**選擇演算法的階段**之中，人工智慧的建置者，應揭露所使用的演算法類型。不同的演算法表彰了不同事件（資料）之間的關聯性。不同的演算法影響了人工智慧在表現上不同的可理解性。不同演算法中可理解性程度的不同，在不同領域中尤其有不同的容許性程度。在本文前述的推論之中，既然自動化決定並沒有不提供理由說明決策產生的特殊原因下，在必須提供理由的領域而自動化決定無法提供理由時，決定者即應該拒絕使用特定類型的人工智慧，以符合法治原則的要求。[65]

在**人工智慧使用的過程**之中，決策者有一般性的說明義務，說明為何特定任務的執行之中，具有使用人工智慧的理由；在特定種類任務之中具有使用人工智慧的理由時，決策者具有個別性的說明義務，說明特定個案是否是該人工智慧能夠處理的個案，是否有特定應納入而未納入考量的資訊，以釐

[62] *See* Cobbe, Lee & Singh, *supra* note 60, at 604-605.

[63] *See* Oswald, *supra* note 29, at 10-14.

[64] *See* KATE CRAWFORD, ATLAS OF AI: POWER, POLITICS, AND THE PLANETARY COSTS OF ARTIFICIAL INTELLIGENCE 123-150 (1 ed. 2021).

[65] *See* Ananny & Crawford, supra note 47, at 985.

doi:10.6342/NTU202103631

清該人工智慧是否能適用於該個案之中。而這個階段的說明義務，尤其以前述設計資料的揭露為基礎，方能夠有效的落實。

由於上述資訊的揭露，在不同的決策領域中，皆具有不同的重要性，因此在立法框架的配套措施上，除了以透明為核心的規範內容外，也必須搭配「人工智慧影響評估」的機制，判斷各個人工智慧使用的脈絡的影響可能性與強烈程度，並且引入相關技術的「專家」，協助人工智慧的使用者作成更細緻的決定。對於透過立法者所制定的這些框架，將會形塑未來對於人工智慧的內涵，如果無法有效的揭露上述資訊的人工智慧系統，在法治原則的要求下，就會排除於公部門的使用，因為無法達成課責目的的工具，則無法在公部門中使用。

上述的框架性規範要求能夠透過一般法制框架的建立，引導公部門所制定的人工智慧的內涵。然而，現實中因為預算的限制以及能力的缺乏，公部門比較可能是透過採購的方式取得私部門製造的人工智慧系統。在這樣的情況之中，除了規範性的要求之外，公部門能夠透過政府採購契約中的要求落實前述的框架性要求。[66]雖然政府採購法制的目的主要在於協助政府獲得「俗又大碗」的服務，並落實公平的競爭環境[67]，但政府採購的法制中也不乏透過對於締結契約廠商的特殊義務要求，以落實特定的公益目的的規範。[68]在人工智慧系統的採購上，尤其應透過契約條款要求私部門揭露上述的資訊，以達成公部門的法治原則要求。透過採購契約條款要求落實上述的揭露義務，不僅較立法框架的建立具有能動性外，公部門的契約條款措施能夠間接的影響企業的產品生態，大範圍的影響未來人工智慧的生產鏈，有效的落實人工智慧的有效治理。[69、70]

---

[66] *See* Cary Coglianese & Erik Lampmann, *Contracting for Algorithmic Accountability*, 6 ADMIN. L. REV. ACC. 175 (2021).

[67] *See* Mulligan & Bamberger, *supra* note 58, at 796-797.

[68] 例如我國政府採購法第98條即課予義務，要求特定規模以上的得標廠商必須聘用一定比例的身心障礙者以及原住民，以保障其權益。

[69] Coglianese & Lampmann, *supra* note 66, at 181-184.

[70] 但在涉及要求揭露私部門所致造的人工智慧系統的內涵時，強烈的反對的理由就是「營業秘密」侵害的主張。就此主張而言，必須要思考的點是前述本文所要求的若干揭露內涵是否涉及營業秘密，以及若涉及營業秘密，兩者利益互相權衡之結果為何。政府方的締約者尤其應思考此因素，以具體的形塑必須揭露的內容。本文並不對此加以深入。

doi:10.6342/NTU202103631

一言以蔽之，透過以法規或者是契約條款揭露人工智慧建置與設計階段的資訊，使用者能夠瞭解自動化決定在特定任務之中的表現能力後，再進一步決定在特定任務之中，是否接受或者拒絕機器所作成的建議或決定，以落實法治原則人為介入下以「人」作為課責對象的要求。在實踐上，法治原則則會著重於判斷作成採納特定自動化決定的「理由」是否足夠，是否有判斷不足之處，以確保人類基於自動化決定的輔助所作成的決定，符合法律的要求。過去，由於單一人工智慧系統的建置，通常由不同的角色介入加以設計，不同的角色皆不諳個別的建置對於整體系統最終的影響，形成了課責的鴻溝。[71]而設計系統的透明要求，則補足了這個鴻溝，建立了課責的鏈結，使法治原則的理由要求能夠深入各個形成人工智慧系統的不同角色之中，達到拘束的效果。

***

總結本章的說明，本文從法律續造在法治原則中的地位出發，說明法律適用者為了追求個案正義，而必須從事法律續造的可能行，並且證成了自動化決定的直接使用，無法符合法治原則的評價性結果。在管制手段上，本文提出「人為介入」的管制方式、說明此種手段的正當性、釐清這種手段於現實中的操作、區分法治原則對於不同領域的要求，並且提出更細部人為介入的法制建議。最後，為確保人為介入在現實中有實踐的可能，本文主張透過「透明」法制框架的設計，要求人工智慧的製造者必須揭露設計階段之中的重要資訊予人工智慧的使用者，使用者能夠透過這些資訊判斷是否在個案中使用人工智慧，以有效的落實人為介入的要求。一言以蔽之，本文分別說明人為介入在憲法法治原則上的正當性，以及在法制規範面上提出具體主張，最終並證成：人為介入作為公部門正當使用人工智慧的必要條件。

[71] *See* Filippo Santoni de Sio & Giulio Mecacci, *Four Responsibility Gaps with Artificial Intelligence: Why they Matter and How to Address them*, PHILOS. TECHNOL. 1, 4-7 (forthcoming) (2021).

doi:10.6342/NTU202103631

第五章

結論

5.1 論證總結

本文基於人類與人工智慧在表現上有所差距為問題意識，釐清公部門在何種條件下能夠正當使用人工智慧，從事適用法律、作成決定的任務。本論文認為：「人為介入」是公部門使用人工智慧從事適用法律、作成決定的必要條件。在論證上，本文首先說明「人為介入」這個主張的詳細內涵（第二章），在分析人類與人工智慧在適用法律上的差異（第三章）以及這些差異在憲法法治原則的評價後，最終證成本文主張的「人為介入」的管制手段具有正當性，以及提出落實人為介入的法制框架（第四章）。

詳細的來說，本文於第二章中以歐盟一般資料保護規則中的「人為介入」作為例子，說明人為介入的實際內涵。歐盟法透過細緻並有層次的架構落實自動化決定的管制目的。蘊藏於這套管制架構背後的一貫主張就是落實「人為介入」的要求。人為介入的事實判斷是：人為決定與自動化決定有所不同、人為決定能夠避免自動化決定的傷害，以及反對全面性的自動化決定。

在釐清人為介入的內涵後，本文進一步分析這種管制模式是否具有正當性。因此，本文於第三章分析人類與人工智慧在進行適用法律、作成決定任務上的差距。本文指出人工智慧的兩種能力，「模仿」以及「發現洞見」。基於單純「資料關聯性」，在法律適用過程中的不可援用，本文鎖定人工智慧的模仿能力，分析其與人類適用法律的差異。本文指出人工智慧與人類進行法律適用時的差異在於，人工智慧沒有辦法對於新事實進行法律適用，以及進行法律續造。

奠基於第三章的這兩種適用法律結果上，本文於第四章從法治原則出發，檢討自動化決定相關特色在憲法框架下的評價結果。本文從法治原則與法律續造的關係出發，推導出公部門在從事法律適用任務中，必須從事法律續造的必要。在釐清人工智慧在憲法上的評價之後，本文指出「人為介入」是一個能夠使人工智慧的使用符合憲法要求的正當性條件，並且進一步的說明人

doi:10.6342/NTU202103631

為介入在法治原則的要求下，應該如何設計人為介入的內涵，以避免現實中的管制目的無法達成。為確保人工智慧的使用以及人為介入的落實，本文主張建構以人工智慧的設計資訊為揭露對象的管制框架，由決策者正視特定人工智慧的優點以及問題，並進一步決定是否在個案之中使用人為介入，以達到法治原則下的課責要求。

doi:10.6342/NTU202103631

5.2　設例解析

在提出了本文的主張，並且回應了本文的研究問題後，這個部分回到本論文一開始所提出的案例，並加以解析。

本案例之中，行政部門以及司法部門分別透過自動化決定來推行國家任務，而這些任務都是適用法律的任務。金管會必須適用證券交易法，判斷市場中是否有違反法律的交易行為；而司法則是就刑法的個別規定判斷個案是否成罪。

問題一與問題二分別涉及兩個不同的問題。問題一涉及的是人工智慧在公部門的使用下，是否會違反法治原則；問題二則涉及若要使用人工智慧，應搭配何種管制措施能有效避免自動化決定違反法治原則。問題一的爭點在，人工智慧的本質是否與憲法框架的內涵相符；而問題二則是涉及透過何種手段，能夠有效的避免人工智慧與憲法框架的齟齬，同時又享受人工智慧的益處。

問題一中分別涉及了金管會使用自動化決定作成行政處分，及法院透過自動化決定作成個案是否成罪認定的兩種國家行為。這兩種國家行為的差異在於前者是行政機關對於個案進行認識用法，後者則是司法機關對於個案是否符合刑事規定進行判斷。

依據本文的論證，人工智慧的使用，將造成金管會作成的行政處分，與法院作成個案成罪與否的決定，無法依據個案的背景條件差異，從事法律的續造。雖然金管會是行政機關，但行政機關在適用法律時，也有必須從事法律續造的義務。此外，法院對於個案是否成罪的認定，尤其必須確保個案的決定。因此，可以得知金管會以及法院透過自動化決定的方式，對於當事人產生影響，與法治原則的要求有所不符。

在問題二「管制手段」的問題上，既然直接使用「自動化決定」將與憲法法治原則有所違背，若仍欲享受人工智慧所帶來的利益，就是採取「人為介入」的方式，將人工智慧在決定形成的過程中置於輔助、備位的角色。

在人為介入的設計上，可以分為兩種類型，一種是輔助人為決定型，另外一種是人為事後介入保留型。這兩種類型的區分標準，尤其應該視個案之

doi:10.6342/NTU202103631

中人工智慧所應用的領域以及所適用的法律要件進行區分。在領域上，基於程序法治原則對於司法制度的特殊要求，司法制度之中尤其必須採取輔助型的人工智慧，必須由人類作成每一個發生法律效果的決定，確保司法作成個案的決定。因此在法院運用人工智慧作成個案是否成罪的認定情形中，僅能採取以「輔助型」的方式加以使用，不得採取「人為事後介入保留」的規範模式。在金管會作成的行政處分中，由於具有大批量的性質，為了效率的促成，本文認為可以採取「人為事後介入保留」的人工智慧應用方式。

此外，人為介入的理念是，人類仍為法治原則課責的對象，因此人類決策者必須確實的評估個案之中是否受到人工智慧的輔助。在這樣的要求下，以「設計階段中的重要資訊」作為揭露對象的透明法制框架即不可或缺。人類決策者必須透過觀察特定人工智慧系統整體的設計理念、細部的形成方式，以評估個案中是否能夠接納、或者拒絕人工智慧的輔助結果，以落實「人為介入」的規範性要求。

doi:10.6342/NTU202103631

5.3　未來展望

人工智慧的開發、研究與使用，在國際的討論上正熱熱烈烈的進行中。但對於這種技術是否能夠使用、使用範圍多大、應如何使用，我們都應該要先仔細的思考。在人工智慧的管制條件上，我國與外國最大的差異在於規範的出發點有所不同。在發展以及應用上，外國的情形是技術的發展快於法律的管制手段，因此相關管制手段皆要屈就於現行的發展，就當下的技術問題發展出當下的管制框架；相較於此，我國在技術發展的進度上起步較晚，但這也意味著我國在法制規範設計上能夠掌握先機，為未來的技術形塑發展方向。因為這個現實上的差距，相關外國法制上管制目的與管制結構的主張，在挪用於我國的時候，尤其必須關注這個特點，而進行本土化的調整。以現行人工智慧所造成的問題為例，由於人工智慧在國外的使用已經是現在進行式，因此在管制手段上尤其著重於人工智慧所造成的傷害辨認以及事後救濟途徑的可行性；相較於此，若我國在規範上僅亦步亦趨的著眼於外國的相關討論，可能就喪失了另一波形塑技術內涵的先機。換言之，與其學習國外針對傷害辨認以及事後救濟進行討論，我國因為技術發展正處於起步階段，在規範面上更應該從源頭下手，透過管制框架的事先設定，避免這些問題的產生。

更具體的來說，本文所提出的規範性與框架性主張就是以這種背景事實為基礎。詳言之，人為介入的規範性要求，就是筆者在綜合考量人工智慧所可能帶來的利益以及傷害後，對於人工智慧的應用所劃定的界線，一方面容許人工智慧輔助人類進行決定，一方面禁止人工智慧完全取代人類進行適用法律、作成決定的任務。此外，為了使人為介入的規範性要求在人工智慧的時代下並不只是一個學術上的主張，本文透過「透明」的框架性創設，使人工智慧輔助人類決定在現實上有實踐的可能。透明框架的創設除了能事先的形塑未來的技術現實外，另外一個優點是在最小影響技術發展界線的前提上，落實課責的要求，將課責的責任轉交於公部門中的人類。筆者相信本文提出的「人為介入」搭配「設計內涵透明」的法制，在理論上以及現實上都是符合臺灣現實科學、技術發展與法規狀態的適合措施。

doi:10.6342/NTU202103631

在公部門的使用容許性評價上，本文主要從憲法的法治原則出發加以論證。但是人工智慧在公部門的使用，並不是僅對於國家權力的行使的基礎有所影響，甚至可能侵害個人的基本權利。例如自動化決定利用非法律所保障的特徵進行決定，並且形成無法確定歧視因素的推論結果時，平等權應如何有效的回應自動化決定所造成的影響。[1]更嚴重的一個問題則是，廣泛使用人工智慧對於個人以及社會所造成的影響。詳言之，人工智慧透過資料對於個人進行評價，個人因為擔心受到不利對待，而有意識的改變行為，而受改變的行為被蒐集、形成資料，再度進入自動化決定的資料形成下一次決定基礎的循環過程中。且人工智慧本就有無法打破常規的特色。如此的循環，對於社會環境所形塑的人格與自主，將受到何種程度的影響，如何避免人的圖像的單一化、避免人類往機器人的圖像邁進也是本文並未處理的困難問題。[2]

最後，本文處理的問題是立基於現行的技術條件，以第二波人工智慧與人類在任務執行的效能上仍有所不同為問題意識，而這也是學界目前的共識。但在面對模仿人類的技術，我們或許無可避免的是，終究必須面對人類與人工智慧在任務執行上完全沒有差異、迎接通用人工智慧到來的那一天。我們如何在這樣的社會中保持「人類的價值」同時與技術共存，或許現在是時候開始思考相關的問題了！

[1] *See*, *e.g.*, Alexander Tischbirek, *Artificial Intelligence and Discrimination: Discriminating Against Discriminatory Systems*, *in* REGULATING ARTIFICIAL INTELLIGENCE 103-121 (Thomas Wischmeyer & Timo Rademacher eds., 2020); Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671 (2016); Aziz Z. Huq, *Racial Equity in Algorithmic Criminal Justice*, 68 DUKE L.J. 1043 (2018); Sandra G. Mayson, *Bias In, Bias Out*, 128 YALE L. J. 2218 (2019); Sandra Wachter, *Affinity Profiling and Discrimination by Association in Online Behaviorial Advertising*, 35 BERKELEY TECH. L. J. 367 (2020); Frederik J. Zuiderveen Borgesius, *Strengthening legal protection against discrimination by algorithms and artificial intelligence*, 24 INT'L J. HUMAN RIGHTS 1572, 1584 (2020).

[2] *See*, *e.g.*, JULIE E. COHEN, CONFIGURING THE NETWORKED SELF: LAW CODE AND THE PLAY OF EVERYDAY PRACTICE (1 ed. 2012). Julie E. Cohen, *What Privacy is for*, 126 HARV. L. REV. 1904, 1911 (2013); Mireille Hildebrandt, *Privacy as Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning*, 20 THEORETICAL INQUIRIES L. 83 (2019).

doi:10.6342/NTU202103631

參考文獻

一、中文文獻

Greco, Luís（著）、鍾宏彬（譯）（2021），〈沒有法官責任的法官權力：為什麼不許有機器人法官〉，《月旦法學雜誌》，315期，頁170-195。

Larenz, Karl（著），陳愛娥（譯）（2013），《法學方法論》，初版九刷，臺北：五南。

Mayer-Schönberger, Viktor（著），林俊宏（譯）（2015），《大數據：隱私篇》，1，臺北：天下。

王怡蘋（2020），〈人工智慧創作與著作權之相關問題〉，收於：李建良（編），《法律思維與制度的智慧轉型》，頁563-598，臺北：

王紀軒（2019），〈人工智慧於司法實務的應用〉，《月旦法學雜誌》，293期，頁93-114。

王鵬翔（2004），〈目的性限縮之論證結構〉，《月旦民商法雜誌》，4期，頁16-29。

--------（2005），〈論基本權的規範結構〉，《臺大法學論叢》，34卷2期，頁1-60。

--------（2005），〈論涵攝的邏輯結構—兼評Larenz的類型理論〉，《成大法學》，9期，頁1-45。

--------（2007），〈規則、原則與法律說理〉，《月旦法學教室》，53期，頁74-83。

--------（2008），〈規則是法律推理的排它性理由嗎？〉，收於：王鵬翔（編），《2008法律思想與社會變遷》，頁345-386，臺北：中央研究院法律研究所籌備處。

王鵬翔、張永健（2015），〈經驗面向的規範意義－論實證研究在法學中的角色〉，《中研院法學期刊》，17期，頁205-294。

何之行、廖貞（2020），〈AI個資爭議在英國與歐盟之經驗－以Google DeepMind一案為例〉，收於：李建良（編），《法律思維與制度的智慧轉型》，頁333-383，臺北：元照。

吳全峰（2020），〈初探人工智慧與生命倫理之關係〉，收於：李建良（編），

doi:10.6342/NTU202103631

《法律思維與制度的智慧轉型》，頁167-222，臺北：元照。

吳從周（2018），〈初探AI的民事責任－聚焦反思臺灣之實務見解〉，收於：劉靜怡（編），《人工智慧相關法律議題芻議》，頁87-117，臺北：元照。

李建良（2020），〈人工智慧與法學變遷－法律人面對科技的反思〉，收於：李建良（編），《法律思維與制度的智慧轉型》，頁1-87，臺北：元照。

--------（2018），〈法學方法與基本權解釋方法導論〉，《人文及社會科學集刊》，30卷2期，頁237-277。

李榮耕（2018），〈初探刑事程序法的人工智慧應用－以犯罪熱區為例〉，收於：劉靜怡（編），《人工智慧相關法律議題芻議》，頁118-148，臺北：元照。

沈宗倫（2018），〈人工智慧科技與智慧財產法治的交會與調和－以著作權法與專立法之權利歸屬為中心〉，收於：劉靜怡（編），《人工智慧相關法律議題芻議》，頁177-214，臺北：元照。

--------（2020），〈人工智慧科技對於專利侵權法制的衝擊與因應之道〉，收於：李建良（編），《法律思維與制度的智慧轉型》，頁523-562，臺北：元照。

林子儀（2015），〈隱私權法制的新議題：監控與隱私自我管理〉，收於：國立臺灣大學法律學院（編），《第五屆馬漢寶講座論文彙編》，頁65-115，臺北：財團法人馬氏思上文教基金會。

林勤富（2018），〈人工智慧法律議題初探〉，《月旦法學雜誌》，274期，頁195-215。

林勤富、李怡俐（2020），〈人工智慧時代下的國際人權法－規範與制度的韌性探索與再建構〉，收於：李建良（編），《法律思維與制度的智慧轉型》，頁413-465，臺北：元照。

邱文聰（2018），〈初探人工智慧中的個資保護發展趨勢與潛在的反歧視難題〉，收於：劉靜怡（編），《人工智慧相關法律議題芻議》，頁145-175，臺北：元照。

--------（2020），〈第二波人工智慧知識學習與生產對法學的挑戰—資訊、科技與社會研究及法學的對話〉，收於：李建良（編），《法律思維與制度的智慧轉型》，頁137-168，臺北：元照。

--------（2021），〈亦步亦趨的模仿還是超前部署的控制？AI的兩種能力和它們帶

doi:10.6342/NTU202103631

來的挑戰〉，作者手稿。即將出版，收於：《智慧新世界？人文社科學人眼中的AI》，新竹：國立清華大學。

洪德欽（2015），〈歐盟法的淵源〉，收於：洪德欽、陳淳文（編），《歐盟法之基礎原則與實務發展（上）》，頁1-56，臺北：國立臺灣大學出版中心。

張嘉尹（2019），〈憲法解釋作為法律續造——個方法論的反思〉，《中原財經法學》，43期，頁1-38。

張陳弘、莊植寧（2019），《新時代之個人資料保護法制：歐盟GDPR與臺灣個人資料保護法的比較說明》，臺北：新學林。

張嘉尹（2002），〈法律原則、法律體系與法概念論－Robert Alexy法律原則理論初探〉，《輔仁法學》，24期，頁1-48。

陳弘儒（2020），〈初探目的解釋在法律人工智慧之運用可能〉，《歐美研究》，50卷2期，頁293-347。

--------（2020），〈初探目的解釋在法律人工智慧系統之運用可能〉，收於：李建良（編），《法律思維與制度的思維轉型》，頁227-302，臺北：元照。

陳柏良（2020），〈AI時代之分裂社會與民主－以美國法之表意自由與觀念市場自由競爭理論為中心〉，收於：李建良（編），《法律思維與制度的智慧轉型》，頁301-331，臺北：元照。

陳顯武（2005），〈論法學上規則與原則的區分－由非單調邏輯之觀點出發〉，《臺大法學論叢》，34卷1期，頁1-45。

黃居正（2018），〈與人工智慧相關的國際法議題－從國際人道法到生命體法〉，收於：劉靜怡（編），《人工智慧相關法律議題芻議》，頁215-241，臺北：元照。

黃舒芃（2013），〈法律保留原則在德國法秩序下的意涵與特徵〉，收於：黃舒芃（編），《民主國家的憲法及其守護者》，頁7-53，臺北：元照。

黃詩淳、邵軒磊（2017），〈運用機器學習預測法院裁判—法資訊學之實踐〉，《月旦法學雜誌》，270期，頁86-96。

--------（2019），〈人工智慧與法律資料分析之方法與應用：以單獨親權酌定裁判的預測模型為例〉，《臺大法學論叢》，48卷4期，頁2023-2073。

--------（2020），〈以人工智慧讀取親權酌定裁判文本：自然語言與文字探勘之實踐〉，《臺大法學論叢》，49卷1期，頁195-224。

doi:10.6342/NTU202103631

黃銘傑（2019），〈人工智慧發展對法律及法律人的影響 〉，《月旦法學教室》，200期，頁51-54。

楊岳平（2020），〈人工智慧時代下的金融監理議題－以理財機器人監理為例〉，收於：李建良（編），《法律思維與制度的智慧轉型》，頁467-500，臺北：元照。

雷磊（2009），〈法律規範的同位階衝突及解決：以法律規則與法律原則的關係為出發點〉，《臺大法學論叢》，38卷4期，頁1-66。

劉靜怡（2018），〈人工智慧潛在倫理與法律議題鳥瞰與初步分析〉，收於：劉靜怡（編），《人工智慧相關法律議題芻議》，頁1-45，臺北：元照。

--------（2020），〈人工智慧時代的法學研究路徑初探〉，收於：李建良（編），《法律思維與制度的智慧轉型》，頁91-134，臺北：元照。

--------（2021），〈科技正當法律程序的憲法意涵—美國判決與學說發展的檢視〉，收錄於氏著：《網路時代的隱私保護困境》，頁356-411，臺北：元照。

謝碩駿（2021），〈論全自動作成之行政處分〉，收於：黃丞儀（編），《2017行政管制與行政爭訟》，先期出版，頁1-86，臺北：中央研究院法律研究所。http://publication.iias.sinica.edu.tw/92015091.pdf

顏厥安（1998），〈法、理性與論證〉，收於：顏厥安（編），《法與實踐理性》，頁95-212，臺北：允晨。

--------（1998），〈法與道德－由一個法哲學的核心問題檢討德國戰後法思想的發展〉，收於：顏厥安（編），《法與理性實踐》，頁31-76，臺北：允晨叢刊。

--------（2002），〈規則、理性與法治〉，《臺大法學論叢》，31卷2期，頁1-58。

--------（2018），〈人之苦難，機器恩典必看顧安慰－人工智慧、心靈與演算法社會〉，收於：劉靜怡（編），《人工智慧相關法律議題芻議》，頁47-85，臺北：元照。

二、英文文獻

Alarie, B. (2016). The path of the law: Towards legal singularity. *University of Toronto Law*

doi:10.6342/NTU202103631

Journal, 66(4), 443-455. https://doi.org/10.3138/UTLJ.4008

Alderucci, D. (2020). The Automation of Legal Reasoning: Customized AI Techniques for the Patent Field. *Duquesne Law Review, 58*(1), 50-81.

Alexy, R. (1978). A Theory of Legal Arguementation. Translated by R. Adler and N. MacCormick. U.S.: Oxford University Press.

-------- (1994). A Theory of Constitutional Rights. Translated by J. Rivers (2nd. ed). New York: Oxford University Press.

Alpaydin, E. (2014). *Introduction to Machine Learning* (3rd. ed). Massachusetts: The MIT Press.

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society, 20*(3), 973-989. https://doi.org/10.1177/1461444816676645

van Andel, P., & Bourcier, D. (2002). Serendipity and Abduction in Proofs, Presumptions and Emerging Laws. In M. MacCrimmon & P. Tillers (Eds.), *The Dynamics of Judicial Proof: Computation, Logic, and Common Sense* (pp.273-286). Heidelberg: Physica.

Appel, S. M., & Coglianese, C. (2020). Algorithmic Governance and Administrative Law. In W. Barfield (Ed.), *The Cambridge Handbook of the Law of Algorithms* (pp. 162-181). New York: Cambridge University Press.

Article 29 Data Working Party. (2018). Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679. https://ec.europa.eu/newsroom/article29/redirection/document/49826

-------- (2018). Guidelines on Transparency under Regulation 2016/679. https://ec.europa.eu/newsroom/article29/redirection/document/51025

Ashley, K. D. (2017). *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age* (1 ed.). New York: Cambridge University Press.

Barocas, S., & Selbst, Andrew D. (2016). Big Data's Disparate Impact. *California Law Review, 104*, 671-732.

Bayamlıoğlu, E. (2018). Contesting Automated Decisions. *European Data Protection Law Review, 4*, 433-446.

Berman, E. (2018). A Government of Laws and not of Machines. *Boston College Law Review, 87*, 1277-1355.

Bermejo, J. M. P. (2012). Principles, Comflicts, and Defeats: An Approach from a Coherentist Theroy. In J. F. Beltrán and G. B. Ratti (Eds.), *The Logic of Legal Requirements: Essays on Defeasibility* (pp. 288-308). U.K.: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199661640.003.0018

doi:10.6342/NTU202103631

Berryhill, J., Heang, K. K., Clogher, R., & McBride, K. (2019). Hello, World: Artificial intelligence and its use in the public sector (OECD Working Papers on Public Governance), available at: https://www.oecd-ilibrary.org/governance/hello-world_726fd39d-en [ogy.de/v9ie].

Binns, R. (2017). Algorithmic Accountability and Public Reason. *Philosophy & Technology, 31*(4), 543-556. https://doi.org/10.1007/s13347-017-0263-5

-------- (2021). Human Judgment in algorithmic loops: Individual justice and automated decision-making. *Regulation & Governance*. https://doi.org/10.1111/rego.12358

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning* (1 ed.). U.K.: Springer.

Bloch-Wehba, H. (2021). Transparency's AI Problem. *Knight First Amendment Institute and Law and Political Economy Project's Data & Democracy Essay Series. Texas A&M University School of Law Legal Studies Research Paper No. 21-13*. Available at: https://knightcolumbia.org/content/transparencys-ai-problem [ogy.de/cuq1].

Boden, M. A. (2018). Artificial Intelligence: A Very Short Introduction. U.K.: Oxford University Press.

Borgesius, F. J. Z. (2020). Strengthening legal protection against discrimination by algorithms and artificial intelligence. *The International Journal of Human Rights, 24*, 1572-1593. https://doi.org/10.1080/13642987.2020.1743976

boyd, d., & Crawford, K. (2012). Critical Questions for Big Data: Provocations for A Cultural, Technological, and Scholarly Phenomenon. *Information, Communication & Society, 15*, 662-679. https://doi.org/10.1080/1369118X.2012.678878

Branting, L. K. (2017). Data-centric and Logic-based Models for automated legal problem solving. *Artificial Intelligence and Law, 25*, 5-27. https://doi.org/10.1007/s10506-017-9193-x

Brennan-Marquez, K., & Henderson, S. E. (2019). Artificial Intelligence and Role-Reversible Judgement. *The Journal of Criminal Law & Criminology, 109*(2), 137-164.

Brennan-Marquez, K., Susser, D., & Levy, K. (2019). Strange Loops: Apparent versus Actual Human Involvement in Automated Decision-Making. *Berkeley Technology Law Journal, 34*, 745-771.

Brennan-Marquez, K. (2017). "Plausible Cause": Explanatory Standards in the Age of Powerful Machines. *Vanderbilt Law Review, 70*(4), 1249-1301.

Brkan, M., & Bonnet, G. (2020). Legal and Technical Feasibility of the GDPR's Quest for Explanation of Algorithmic Decisions: of Black Boxes, White Boxes and Fata Morganas. *European Journal of Risk Regulation, 11*(1), 18-50.

doi:10.6342/NTU202103631

https://doi.org/10.1017/err.2020.10

Brkan, M. (2019). Do algorithms rule the world? Algorithmic decision-making and data protection in the framework of the GDPR and beyond. *International Journal of Law and Information Technology, 27*, 91-121. https://doi.org/10.1093/ijlit/eay017

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society, 3*(1), 16. https://doi.org/10.1177/2053951715622512

Bygrave, L. A. (2019). Minding the Machine v2.0: The EU General Data Protection Regulation and Automated Decision Making. In K. Yeung and M. Lodge (Eds.), *Algorithmic Regulation* (pp. 248-262). U.K.: Oxford University Press. https://doi.org/10.1093/oso/9780198838494.003.0011

-------- (2020). Article 22 Automated individual decision-making, including profiling. In C. Kuner, L. A. Bygrave and C. Docksey (Eds.), *The EU General Data Protection Regulation (GDPR): A Commentart* (1 ed., pp. 522-542). U.K.: Oxford University Press. https://doi.org/10.1093/oso/9780198826491.003.0055

Calo, R., & Citron, D. K. (2021). The Automated Administrative State: A Crisis of Legitimacy. *Emory Law Review, 70*, 797-845.

Carlos, J. B. (2001). Why is Legal Reasoning Defeasible? In A. Soeteman (Ed.), *Pluralism and Law* (pp. 327-346). Switzerland: Springer. https://doi.org/10.1007/978-94-017-2702-0_17

Casey, A., & Niblett, A. (2015). The Death of Rules and Standards. *Indiana Law Journal, 92*, 1401-1448.

Casey, B., Farhangi, A., & Vogl, R. (2019). Rethinking Explainable Machines: The GDPR's 'Right to Explanation' Debate and the Rise of Algorithmic Audits in Enterprise. *Berkeley Technology Law Journal, 34*, 145-189.

Chander, A. (2016). The Racist Algorithm. *Michigan Law Review, 115*, 1023-1046.

Citron, D. K., & Pasquale, F. (2014). The Scored Society: Due Process For Automated Predictions. *Washington Law Review, 89*, 1-33.

Citron, D. K. (2007). Technological Due Process. *Washington University Law Review, 85*, 1249-1314.

Cobbe, J., Lee, M. S. A., & Singh, J. (2021). Reviewable Automated Decision-Making: A Framework for Accountable Algorithmic Systems. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*, 598-609. https://doi.org/10.1145/3442188.344592

Cobbe, J. (2019). Administrative Law and the Machines of Government: Judicial Review of

doi:10.6342/NTU202103631

Automated Public Sector Decision-making. *Legal Studies, 39*, 636-655. https://doi.org/10.1017/lst.2019.9

Coglianese, C., & Lampmann, E. (2021). Contracting for Algorithmic Accountability. *Administrative Law Review Accord, 6*, 175-199.

Coglianese, C., & Lehr, D. (2017). Regulating by Robot: Administrative Decision Making in the Machine-Learning Era. *The Georgetown Law Journal, 105*, 1147-1223.

-------- (2019). Transparency and Algorithmic Governance. *Administrative Law Review, 71*, 1-56.

Cohen, J. E. (2012). *Configuring the Networked Self: Law Code and the Play of Everyday Practice* (1 ed.). New Heaven: Yale University Press.

-------- (2013). What Privacy is for. *Harvard Law Review, 126*, 1904-1933.

-------- (2019). *Between Truth and Power: The Legal Constructions of Information Capitalism* (1 ed.). U.S.: Oxford University Press.

Cohen, M. (2010). The Rule of Law as the Rule of Reasons. *Archiv für Rechts- und Sozialphilosophie 96*(1), 1-16.

Comment. (2017). State v. Loomis: Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing. *Harvard Law Review, 130*, 1530-1537.

Crawford, K., & Schultz, J. (2014). Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms. *Boston College Law Review, 55*, 93-128.

Crawford, K. (2021). *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (1 ed.). New Haven: Yale University Press.

Crootof, R. (2019). "Cyborg Justice" and the Risk of Technological-Legal Lock-In. *Columbia Law Review Forum, 119*, 233-251.

Cukier, K., Mayer-Schönberger, V., & de Vericourt, F. (2021). *Framers: Human Advantage in an Age of Technology and Turmoil* (1 ed.). U.S.: Dutton.

Danziger, S., Levav, J., & Avnaim-Pesso, L. (2011). Extraneous Factors in Judicial Decisions. *Proceedings of the National Academy of Sciences of the United States of America, 108,* 6889-6892.

Deakin, S., & Markou, C. (2020). Ex Machina Lex: The Limits of Legal Computability. In C. Markou and S. Deakin (Eds.), *Is Law Computable? Critical Perspectives on Law and Artificial Intelligence* (pp. 31-66). U.K.: Hart Publishing.

Deeks, A. (2019). The Judicial Demand for Explainable Artificial Intelligence. *Columbia Law Review, 119*, 1829-1850.

-------- (2020). Secret Reason-Giving. *Yale Law Journal, 129*, 612-689.

Diakopoulos, N. (2020). Transparency. In M. D. Dubber, F. Pasquale & S. Das (Eds.), *The*

doi:10.6342/NTU202103631

Oxford Handbook of Ethics of AI (pp. 197-214). New York: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190067397.013.11

Djeffal, C. (2020). Artificial Intelligence and Public Governance: Normative Guidelines for Artificial Intelligence in Government and Public Administration. In T. Wischmeyer and T. Rademacher (Eds.), *Regulating Artificial Intelligence* (pp. 277-293). Switzerland: Springer Nature Switzerland AG. https://doi.org/10.1007/978-3-030-32361-5_12

Dworkin, R. (1977). *Taking Rights Seriously* (1 ed.). Cambridge: Harvard University Press.

Eaglin, J. M. (2017). Constructing Recidivism Risk. *Emory Law Review, 67*, 59-122.

Edwards, L. & Veale, M. (2017). Slave to the Algorithm: Why a Right to an Explanation Is Probably Not the Remedy You Are Looking for. *Duke Law & Technology Review, 16*, 18-84.

Engstrom, D. F., Ho, D. E., Sharkey, C. M., & Cuéllar, M.-F. (2020). *Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies*. Retrieved from https://www.acus.gov/sites/default/files/documents/Government%20by%20Algorithm.pdf [https://ogy.de/b1kr].

Ernst, C. (2020). Artificial Intelligence and Autonomy: Self-Determination in the Age of Automated Systems. In T. Wischmeyer and T. Rademacher (Eds.), *Regulating Artificial Intelligence* (1 ed., pp. 53-73). Springer: Springer International Publishing. https://doi.org/10.1007/978-3-030-32361-5_3

European Commission. (2019). Building Trust in Human-Centric Artificial Intelligence. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52019DC0168 [ogy.de/26vz]

-------- (2020). White Paper on Artificial Intelligence – A European Approach to excellence and trust. https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en [ogy.de/kgps]

Fagen, F., & Levmore, S. (2019). The Impact of Artificial Intelligence on Rules, Standards, and Judicial Discretion. *Southern California Law Review, 93*(1), 1-36.

Fallon, R. H., Jr. (1997). "The Rule of Law" as a Concept in Constitutional Discourse. *Columbia Law Review, 97*, 1-56.

Franklin, S. (2014). History, motivation, and core themes. In K. Frankish and W. M. Ramsey (Eds.), *The Cambridge Handbook of Artifitial Intelligence* (pp. 15-33). U.K.: Cambridge University Press. https://doi.org/10.1017/CBO9781139046855.003

Fuller, L. (1969). The Morality of Law (revised edition). U.S.: Yale University Press.

Gillis, T. B., & Simons, J. (2019). Explanation < Justification: GDPR and the Perils of

doi:10.6342/NTU202103631

Privacy. *Journal of Law and Innovation, 2*, 71-99.

Gillis, T. B., & Spiess, J. L. (2019). Big Data and Discrimination. *The University of Chicago Law Review, 86*, 459-487.

Goldenfein, J. (2019). Algorithmic Transparency and Decision-Making Accountability: Thought for Buying Machine Learning Algorithms. In S. Bluemmel (Ed.), *Closer to the Machine: Technical, Social, and Legal Aspects of AI* (pp. 41-61). Australia: Office of the Victorian Information Commissioner.

Goodman, B., & Flaxman, S. (2017). European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation". *AI Magazine, 38*(3), 50-57. https://doi.org/10.1609/aimag.v38i3.2741

Governmental Accountability Office. (2021). Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities. https://www.gao.gov/assets/gao-21-519sp.pdf [ogy.de/zqze]

Hage, J., & Peczenik, A. (2000). Law, Moral and Defeasibility. *Ratio Juris, 13*(3), 305-325.

Hage, J. (1997). *Reasoning with Rules: An Essay on Legal Reasoning and Its Underlying Logic* (1 ed.). Switzerland: Springer Science.

-------- (2003). Law and Defeasibility. *Artificial Intelligence and Law, 11*, 221-243. https://doi.org/10.1023/B:ARTI.0000046011.13621.08

Hart, H. L. A. (1948-1949). The Ascription of Responsibility and Rights. *Proceedings of the Aristotelian Society, 49*, 171-194.

Hildebrandt, M. (2008). A Vision of Ambient Law. In R. Brownsword and K. Yeung (Eds.), *Regulating Technologies: Legal Futures, Regulatory Frames and Technologocal Fixes* (1 ed., pp. 175-192). U.S.: Hart Publishing. https://doi.org/10.5040/9781472564559.ch-008

-------- (2015). *Smart Technologies and the End(s) of Law* (1 ed.). U.K.: Edward Elgar.

-------- (2019). Privacy as Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning. *Theoretical Inquiries in Law, 20*, 83-121. https://doi.org/10.1515/til-2019-0004

Hong, S.-H. (2020). *Technologies of Speculation: The Limits of Knowledge in a Data-Driven Society* (1 ed.). New York: New York University Press.

Hosanagar, K. (2019). *A Human Guide to Machine Intelligence* (1 ed.). N.Y.: Viking.

Huq, A. Z. (2018). Racial Equity in Algorithmic Criminal Justice. *Duke Law Journal, 68*, 1043-1134.

-------- (2020). A Right to a Human Decision. *Virginia Law Review, 106*, 611-688.

Jones, M. L. (2017). The right to a human in the loop: Political constructions of computer

doi:10.6342/NTU202103631

automation and personhood. *Social Studies of Science, 47*(2), 216-239. https://doi.org/10.1177/0306312717699716.

Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). *Noise* (1 ed). U.S.: William Collins.

Kaminski, M. E., & Malgieri, G. (2021). Algorithmic impact assessments under the GDPR: producing multi-layered explanations. *International Data Privacy Law. 11*, 125-144. https://doi.org/10.1093/idpl/ipaa020.

Kaminski, M. E. (2018). Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability. *Southern California Law Review, 92*, 1529-1616.

-------- (2020). Understanding Transparency in Algorithmic Accountability. In W. Barfield (Ed.), *The Cambridge Handbook of the Law of Algorithms* (pp. 121-138). New York: Cambridge University Press. http://doi.org/10.1017/9781108680844.006

Katyal, S. K. (2019). Private Accountability in the Age of Artificial Intelligence. *UCLA Law Review, 66*, 54-141.

Kennedy, R. (2020). The Rule of Law and Algorithmic Governance. In W. Barfield (Ed.), *The Cambridge Handbook of the Law of Algorithms* (pp. 209-232). New York: Cambridge University Press. https://doi.org/10.1017/9781108680844.012

Kim, P. T. (2017). Data-Driven Discrimination at Work. *William & Mary Law Review, 58*, 857-936.

Klimas, T., & Vaitiukait, J. (2008). The Law of Recitals in European Community Legislation. *ILSA Journal of International & Comparative Law, 15*(1), 61-93.

Kroll, J. A., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2016). Accountable Algorithms. *University of Pennsylvania Law Review, 165*, 633-706.

Larson, E. J. (2021). *The Myth of Artificial Intelligence: Why Computers Can't Think the Way We Do* (1 ed.). U.S.: Harvard University Press.

Launchbury, J. (2017). A DARPA Perspective on Artificial Intelligence. https://www.darpa.mil/attachments/AIFull.pdf [ogy.de/nh5g]

Lehr, D., & Ohm, P. (2017). Playing with the Data: What Legal Scholars Should Learn About Machine Learning. *University of California, Davis Law Review, 51*, 653-717.

Leonelli, S. (2020). Scientific Research and Big Data. In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/sum2020/entries/science-big-data [ogy.de/zbca]

Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2021). Explainable AI: A Review of Machine Learning Interpretability Methods. *Entropy, 23(1)*(18), 1-45. https://doi.org/10.1017/10.3390/e23010018

doi:10.6342/NTU202103631

Liu, H.-W., Lin, C.-F., & Chen, Yu-Jie. (2019). Beyond State v Loomis: artificial intelligence, government algorithmization and accountability. *International Journal of Law and Information Technology, 27*(2), 122-141. http://doi.org/10.1093/ijlit/eaz001

MacCormick, N. (2005). *Rhetoric and the Rule of Law: A Theory of Legal Reasoning* (1 ed.). New York: Oxford University Press.

Malgieri, G., & Comandé, G. (2017). Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation. *International Data Privacy Law, 7*, 243-265. https://doi.org/10.1093/idpl/ipx019

Malgieri, G. (2019). Automated decision-making in the EU Member States: The right to explanation and other "suitable safeguards" in the national legislations. *Computer Law & Security Review, 35*(5). https://doi.org/10.1016/j.clsr.2019.05.002

Mayer-Schonberger, V., & Cukier, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work, and Think* (1 ed.). New York: Houghton Mifflin Harcourt Publishing Company.

Mayson, S. G. (2019). Bias In, Bias Out. *Yale Law Journal, 128*, 2218-2300.

Mazzocchi, F. (2015). Could Big Data be the end of theory in science?, A few remarks on the epistemology of data-driven science. *EMBO Reports, 16*, 1250-1255. https://doi.org/10.15252/embr.201541001

Medvedeva, M., Vols, M., & Wieling, M. (2020). Using machine learning to predict decisions of the European Court of Human Rights. *Artificial Intelligence and Law, 28*, 237-266.

Mendoza, I., & Bygrave, L. A. (2017). The Right Not to be Subject to Automated Decisions Based on Profiling. In T.-E. Synodinou, P. Jougleux, C. Markou and T. Prastitou (Eds.), *EU Internet Law: Regulation and Enforcement* (pp. 77-98): Springer. https://doi.org/10.1007/978-3-319-64955-9_4

Michaels, A. C. (2020). Artificial intelligence, Legal Change, and Separation of Powers. *University of Cincinnati Law Review, 88*, 1083-1104.

Miller, T. (2019). Explanation in Artificial Intelligence: Insights from the Social Sciences. *Artificial Intelligence, 267*, 38. https://doi.org/10.1016/j.artint.2018.07.007

Mormann, F. (2021). Beyond Algorithms: Toward a Normative Theory of Automated Regulation. *Boston College Law Review, 62*, 1-58.

Moses, L. B. (2020). Not a Single Singularity. In S. Deakin and C. Markou (Eds.), *Is Law Computable?: Critical Perspectives on Law and Artificial Intelligence* (pp. 205-222). U.K.: Hart Publishing.

Mulligan, D. K., & Bamberger, K. A. (2019). Procurement As Policy: Administrative Process

doi:10.6342/NTU202103631

for Machine Learning. *Berkeley Technology Law Journal, 34*, 781-858.

Nay, J., & Standburg, K. J. (2021). Generalizability: Machine Learning and humans-in-the-loop. In R. Vogel (Ed.), *Research Handbook on Big Data Law* (pp. 285-303). USA: Edward Elgar Publishing. https://doi.org/10.4337/9781788972826.00020

Organization for Economic Cooperation and Development. (2019). OECD Principles on AI. https://www.oecd.org/going-digital/ai/principles/ [ogy.de/gvhs]

Oswald, M. (2018). Algorithm-assisted decision-making in the public sector: framing the issues using administrative law rules governing discretionary power. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376*(20170359), 1-20. https://doi.org/10.1098/rsta.2017.0359

Pasquale, F. A. (2019). A Rule of Persons, Not Machines: The Limits of Legal Automation. *George Washington Law Review, 87*, 1-55.

-------- (2016). *The Black Box Society: The Secret Algorithms That Control Money and Information* (1 ed.). MA: Harvard University Press.

-------- (2020). *New Laws of Robotics: Defending Human Expertise in the Age of AI* (1 ed.). Cambridge, Massachusetts: Harvard University Press.

Pearl, J., & Mackenzie, D. (2018). *The Book of Why: The New Science of Cause and Effect*. New York: Basic Books.

Personal Data Protection Commission in Singapore. (2020). Model Artificial Intelligence Governance Framework (Second Edition). https://www.pdpc.gov.sg/help-and-resources/2020/01/second-edition-of-model-artificial-intelligence-governance-framework [ogy.de/7bra]

Postman, N. (1993). *Technopoly: The Surrender of Culture to Technology* (1 ed.). New York: Vintage Books.

Prince, A. E. R., & Schwarcz, D. (2020). Proxy Discrimination in the Age of Artificial Intelligence and Big Data. *Iowa Law Review, 105*, 1257-1318.

Proposal for a Regulation of the European Parliament and of the Council laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending certain Union Legislative acts. (2021). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206 [ogy.de/j92v]

Re, R. M., & Solow-Niederman, A. (2019). Developing Artificially Intelligent Justice. *Stanford Technology Law Review, 22*, 242-289.

Remus, D., & Levy, F. (2017). Can Robots Be Lawyers: Computers, Lawyers, and the Practice of Law. *Georgetown Journal of Legal Ethics, 30*(3), 501-558.

doi:10.6342/NTU202103631

Rissland, E. L. (1990). Artificial Intelligence and Law: Stepping Stones to a Model of Legal Reasoning. *Yale Law Journal, 99*(8), 1957-1982.

Russel, S., & Norvig, P. (2021). Artificial Intelligence: A Modern Approach (4 ed.). U.K.: Pearson.

Sancho, D. (2020). Automated Decision-Making under Article 22 GDPR:Towards a More Substantial Regime for Solely Automated Decision-Making. In M. Ebers and S. Navas (Eds.), *Algorithms and Law* (pp. 136-156). U.K.: Cambridge University Press. https://doi.org/10.1017/9781108347846.005

de Sio, F. S., & Mecacci, G. (2021). Four Responsibility Gaps with Artificial Intelligence: Why they Matter and How to Address them. *Philosophy & Technology*, 1-28. https://doi.org/10.1007/s13347-021-00450-x

Sartor, G. (1995). Defeasibility in Legal Reasoning. In Z. Bankowski, I. White and U. Hahn (Eds.), *Informatics and the Foundations of Legal Reasoning* (pp. 119-158). Switzerland: Springer.

Schauer, F. (2012). Is Defeasibility an Essential Property of Law?. In J. F. Beltrán and G. B. Ratti (Eds.), *The Logic of Legal Requirements: Essays on Defeasibility* (pp. 77-88). U.K.: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199661640.003.0005

Selbst, A. D., & Barocas, S. (2018). The Intuitive Appeal of Explainable Machines. *Fordham Law Review, 87*, 1085-1139.

Selbst, A. D., & Powles, J. (2017). Meaningful Information and the Right to Explanation. *International Data Privacy Law, 7*, 233-242. https://doi.org/233-242. 10.1093/idpl/ipx022

Sunstein, C. R. (2021). Governing by Algorithm? No Noise and (Potentially) Less Bias. *Duke Law Journal*, *forthcoming*. Available at: https://ssrn.com/abstract=3925240 [ogy.de/d8t8]

Surden, H. (2014). Machine Learning and Law. *Washington Law Review, 89*, 87-115.

-------- (2019). Artificial Intelligence and Law: An Overview. *Georgia State University Law Review, 35*, 1305-1337.

Tene, O., & Polonets, J. (2013). Judged by the Tin Man: Individual Rights in the Age of Big Data. *Journal on Telecommunications and High Technology Law, 11*, 351-368.

Tischbirek, A. (2020). Artificial Intelligence and Discrimination: Discriminating Against Discriminatory Systems. In T. Wischmeyer and T. Rademacher (Eds.), *Regulating Artificial Intelligence* (1 ed., pp. 103-121). Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-030-32361-5_5

doi:10.6342/NTU202103631

Tosoni, L. (2021). The right to object to automated individual decisions: resolving the ambiguity of Article 22(1) of the General Data Protection Regulation. *International Data Privacy Law, 11*, 145-162. https://doi.org/10.1093/idpl/ipaa024

Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind, 59*(236), 433-460. https://doi.org/10.1093/mind/LIX.236.433

Veale, M., & Edwards, L.. (2018). Clarity, Surprises, and Further Questions in the Article 29 Working Party Draft Guidance on Automated Decision-Making and Profiling. *Computer Law & Security Review, 34*(2), 398-404. https://doi.org/398-404. 10.1016/j.clsr.2017.12.002

Voigt, P., & von dem Bussche, A. (2017). *The EU General Data Protection Regulation (GDPR): A Practical Guide* (1 ed.). Switzerland: Springer International Publishing.

Volokh, E. (2019). Chief Justice Robots. *Duke Law Journal, 68*, 1135-1192.

Vrabec, H. U. (2021). *Data Subject Right under the GDPR* (1 ed.). U.K.: Oxford University Press.

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. *International Data Privacy Law, 7*, 76-99. https://doi.org/10.1093/idpl/ipx005

Wachter, S., Mittelstadt, B., & Russell, C. (2017). Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR. *Harvard Journal of Law & Technology, 31*, 841-888.

Wachter, S. (2020). Affinity Profiling and Discrimination by Association in Online Behaviorial Advertising. *Berkeley Technology Law Journal, 35*, 367-430.

Waldman, A. E. (2019). Power, Process, and Automated Decision-Making. *Fordham Law Review, 88*, 613-632.

Waldron, J. (2011). The Rule of Law and the Importance of Procedure. In J. E. Fleming (Ed.), *Getting to the Rule of Law* (pp. 3-31). New York: New York University Press.

Wilkinson, J. H., III. (1989). The Role of Reason in the Rule of Law. *The University of Chicago Law Review, 56*, 779-809.

Yeung, K. (2018). Algorithmic regulation: A critical interrogation. *Regulation & Governance, 12*(4), 505-523. https://doi.org/505-523. 10.1111/rego.12158

-------- (2019). Why Worry about Decision-Making by Machine? In K. Yeung and M. Lodge (Eds.), *Algorithmic Regulation* (pp. 21-48). New York: Oxford University Press.

Yu, P. K. (2020). Artificial Intelligence, the Law-Machine Interface, and Fair Use Automation. *Alabama Law Review, 72*(1), 187-238.

doi:10.6342/NTU202103631

Zanfir-Fortuna, G. (2020). Article 13 Information to be provided where personal data are collected from the data subject. In C. Kuner, L. A. Bygrave & C. Docksey (Eds.), *The EU General Data Protection Regulation (GDPR): A Commentary* (pp. 413-433). U.K.: Oxford University Press. https://doi.org/10.1093/oso/9780198826491.003.0044

-------- (2020). Article 15 Right of access by the data subject. In C. Kuner, L. A. Bygrave & C. Docksey (Eds.), *The EU General Data Protection Regulation (GDPR): A Commentary* (pp. 449-468). U.K.: Oxford University Press. https://doi.org/10.1093/oso/9780198826491.003.0046

Zarsky, T. Z. (2013). Transparency in Data Mining: From Theory to Practice. In B. Custers, T. Calders, B. Schermer and T. Zarsky (Eds.), *Discrimination and Privacy in the Information Society* (pp. 301-324). Berlin: Springer-Verlag Berlin Heidelberg. https://doi.org/10.1007/978-3-642-30487-3_17

-------- (2013). Transparent Predictions. *University of Illinois Law Review, 2013*, 1503-1570.

三、德文文獻

Wolff, H. A., & Brink, S. (Hrsg.) (2020). *BeckOK Datenschutzrecht* (32. Aulf.). München: C.H.Beck.

Paul, B. P., & Pauly, D. A. (Hrsg.) (2018). *DS-GVO GDSG* (2. Aulf.). München: C.H.Beck.

Kumkar, L. K., & Roth-Isigkeit, D. (2020), Erklärungspflichten bei automatisierten Datenverarbeitungen nach der DSGVO. *JuristenZeitung*, 277-286.

Käde, L., & von Maltzan, S. (2020), Die Erklärbarkeit von Künstlicher Intelligenz (KI): Entmystifizierung der Black Box und Chancen für das Recht. *Computer und Recht*, 66-72.

doi:10.6342/NTU202103631