科技部人文社會科學研究中心

學術研究群成果報告

人工智慧與法律規範學術研究群

學術研究群編號：MOST 107-2420-H-002-007-MY3- SG10908
學術研究群執行期間：109 年 7 月 1 日至 109 年 12 月 31 日
學術研究群召集人：李建良
執行機構及系所：中央研究院法律學研究所

中　華　民　國　109　年　12　月　31　日

補助學術研究群暨經典研讀班結案報告

計畫名稱

人工智慧與法律規範學術研究群

計畫編號：MOST 107-2420-H-002-007-MY3-SG10908
執行期間： 109 年 7 月 1 日至 109 年 12 月 31 日

執行機構及系所：中央研究院法律學研究所

計畫召集人：李建良
計畫成員：李建良、劉靜怡、王怡蘋、沈宗倫、黃詩淳、林建中、吳全峰、邱文聰、林勤富、楊岳平、蘇凱平、鄭瑞健、何之行、陳弘儒、陳柏良
兼任助理：洪于庭

中 華 民 國 109 年 12 月

補助學術研究群暨經典研讀班成果自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現（簡要敘述成果是否具有政策應用參考價值及具影響公共利益之重大發現）或其他有關價值等，作一綜合評估。

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估

   ☑達成目標

   □ 未達成目標（請說明）

   說明：

2. 研究成果在學術期刊發表或申請專利等情形(請於其他欄註明專利及技轉之證號、合約、申請及洽談等詳細資訊)

   論文：□已發表□未發表之文稿 □撰寫中☑無

   專書：☑已出版□尚未出版□撰寫中□無

   其他：

3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（敘述成果所代表之意義、價值、影響或進一步發展之可能性）。

補助學術研究群暨經典研讀班成果彙整表

計畫主持人：	計畫編號：
計畫名稱：	

成果項目			量化	單位	質化 （說明：各成果項目請附佐證資料或細項說明，如期刊名稱、年份、卷期、起訖頁數、證號...等）
國內	學術性論文	期刊論文	0	篇	請附期刊資訊。
		研討會論文	0		
		專書	1	本	請附專書資訊。
		專書論文	0	章	請附專書論文資訊。
		其他	0	篇	
國外	學術性論文	期刊論文	0	篇	請附期刊資訊。
		研討會論文	0		
		專書	0	本	請附專書資訊。
		專書論文	0	章	請附專書論文資訊。
		其他	0	篇	
參與計畫人力	本國籍	教授	13	人次	
		副教授	6		
		助理教授	9		
		博士後研究員	3		
		專任助理	0		
	非本國籍	教授	0		
		副教授	0		
		助理教授	0		
		博士後研究員	0		
		專任助理	0		
其他成果 (無法以量化表達之成果如辦理學術活動、獲得獎項、重要國際合作、研究成果國際影響力及其他協助產業技術發展之具體效益事項等，請以文字敘述填列。)					

摘要

透過人工智慧改善人類生活品質，為人類政治、經濟、社會與文化模式的重要元素之一。隨著人工智慧的研發進化、運作方式與應用領域的不斷擴增，可以預見人工智慧將對人類社會各種面向與層次造成重大的衝擊，無疑也會帶來或顯或潛的社會問題、倫理議題與法律爭議。**科技發展與法律規範相互碰撞，**牽動的不只是特定的**科學技術領域**，更大程度地可能衝撞到既有的規範板塊，致生法秩序的推移與變革。**法律**人的基本思維是否受到波及？法制運行如何知所因應？法律學門是否在臺灣乃至世界的人工智慧研究圈與學術鏈中扮演關鍵性的角色？如何進行法律學門內部與跨學門之間的整合研究？哪些因人工智慧而起的法律與風險等議題需要研析、探索、防範，降低人工智慧應用的負面衝擊與社會成本？面對以上種種**紛亂複雜的多方法律爭議與研究議題，乃有**以「人工智慧與法律規範」為主題建立法律學術研究群的構想。

本學術研究群的建立與運作，旨在嘗試集臺灣法學界群體之力，以憲法、行政法、刑事法、民事法、商事法、金融法、智財法為經，以法哲學、國際法、區域法、法律史為緯，結合國內關注此一領域且學有專精的法律學者，共同投入人工智慧法律問題的研究，進行廣度與深度兼備的理論探索與實務分析。本學術研究群首先將分別耙梳各法領域與人工智慧的交互關係，依各領域的規範特性，分析人工智慧對各該法領域產生的影響與法律問題，並以法哲學、法律史與國際法/區域法作為橫向的連結，進行貫穿各該法領域的基礎研究。其次，延續基礎研究的成果，擇定若干主題，例如自駕車的法律責任問題，進行深入研究。接著是法律學門自身的融貫研究與跨學門領域的科際整合，前者是憲法、行政法、刑事法、民事法、商事法、金融法、智財法、法哲學/法律史、國際法/區域法之間的問題整合與關聯研究；後者是法律學門與其他領域如資訊科技、哲學、倫理學、社會學等學門之間的問題比較與觀點砥礪。

本學術研究群設定 10 項具前瞻性的子題，分別是：子題 1：人工智慧法律議題的基礎研究與初步分析；子題 2：人工智慧的法哲學與人人類圖像的基本課題；子題 3：人工智慧的憲法課題；子題 4：人工智慧的民事法課題；子題 5：人工智慧的行政法課題；子題 6：人工智慧的刑事法課題；子題 7：人工智慧的智慧財產法課題；子題 8：人工智慧的商事財金法課題；子題 9：人工智慧的區域法課題；子題 10：人工智慧的國際法課題（詳見計畫申請書）。

本學術研究群採定期聚會與不定期座談的方式進行，每月固定聚會一次，計 12 次。定期聚會除依前述規劃的子題進行文獻研讀討論外，將由成員輪流提出報告或邀請相關學者專家演講，同時陸續將活動中演講或報告的相關資料撰寫成論文，發表於期刊或研討會，或集結成專書，作為本學術研究群的計畫成果。第一年自 2018 年 7 月起至 2019 年 6 月底為止，定期舉行 12 次的研究聚會，並於 2019 年

6 月間舉行「人工智慧與法律規範工作坊」，完成第一年計畫，專書《法律思維與制度的智慧轉型：人工智慧與法律規範基礎篇》即將於 2020 年 7 月出版。第二年計畫自 2019 年 7 月開始執行，截至 2020 年 2 月底，已定期舉行 8 次的研究聚會，並預計於 6 月底舉行「《法律思維與制度的智慧轉型》新書發表會」暨「人工智慧政策白皮書工作坊」，以作為本研究群第一年及第二年的計畫成果。

關鍵詞：人工智慧、法律規範、AI

目錄

第一次會議紀錄	
時間	109 年 7 月 16 日（四）
主題	以人工智慧輔助 COVID-19 防治之法律議題初探
講者	吳建昌（臺灣大學法律學院科際整合研究所合聘副教授、臺灣大學醫學院附設醫院精神醫學部主治醫師）

內容摘要

一、人工智慧與大數據之定義與相關倫理原則

（一）、 AI 之定義

AI 主要是電腦系統，可以獲得數據以及詮釋數據並學習，從數據中導引某種結論或結果，再運用在其他或他設定的領域中。另外，機器學習是 AI 的一部份，因為 AI 範圍十分廣闊，機器學習部分是要給他數據，給予 input 再設定 output 的模式，讓機器學習 input 最後要與 output 相對應。另外一種是讓機器找出 input data 彼此間的關聯性，學習一段時間後，一個是設定好數據集讓機器學習，或是不要監督，他也可在沒人監督的情形下進行學習。再者，有關深度學習，是強調基於神經學網絡的演算法，層次可以十分多元，從基礎資訊推到模式的出現。許多 AI 學者會認為深度學習最容易被黑盒子化，因為複雜性高，以至於有些時候難以解釋，故透明性較低。

（二）、 科技部人工智慧科研發展指引：

去年九月科技部提出人工智慧科研發展指引的幾個大原則：

1. 共榮共利(Common Good and Well-being)：科研人員應追求人類、社會、環境間的利益平衡與共同福祉，並致力於多元文化、社會包容、環境永續等，達成保障人類身心健康、創建全體人民利益、總體環境躍升之 AI 社會。
2. 公平性、非歧視性(Fairness and Non-discrimination)：從數據本身或是詮釋的結果都可能會有公平性或歧視性的疑慮。
3. 自主權與控制權(Autonomy and Control)：應用係以輔助人類決策。將來 AI 的功能越來越好時，人類與 AI 可能的共處模式。
4. 安全性(Safety)：我們會希望 AI 是安全的，為合理、善意的使用，構築安全可靠之 AI 環境，故許多人反對 AI 作為武器操作系統。
5. 個人隱私與數據治理(Privacy and Data Governance)：現在一提到資料與個資，大家通常會開始擔心是否會有隱私權之侵害，有無關於人性尊嚴的違反，並會希望有良好的個人資料保護的架構、管理的措施，這兩者是難以分開的
6. 透明性與可追溯性(Transparency and Traceability)：AI 技術需有透明性及可追溯性。關於透明性，如能有某種程度的瞭解，較能夠產生信任。又如能知道從何而來的決策與資料，叫能夠追蹤可以知道最後是哪的問題，故可追溯性會與問責有關。

7. 可解釋性(Explainability)：如 AI 設計本身能夠讓我們知道內在階段性的決策或是結論如何歸納，如能給予以人的觀點作解釋，人類對 AI 會較有信心，與前面的透明性及可追溯性都是有關聯的。
8. 問責與溝通(Accountability and Communication)：最後是希望能有良好的溝通，能將問責系統建立。

如果要讓可解釋性提升，AI 的演算法會受到某種程度的限制，兩者間似乎為 trade-off 的關係。最後可能要考慮，是希望準確性高，或是可解釋性高。

（三）、 大數據定義

大數據的特徵可以包括：數量大、速度快、資料種類及來源多元、有價值、變異性高、虛擬性等。另外，真實性或不確定性也須考量。大數據資料的「巨量」本身具有動態性，有時我們需要強化資料的搜集、分析，連結較大型資料庫的計算力量以及準確度。故有時候巨量本身不需要資料非常大。再者，從大數據中辨認模式，可能幫助提出經濟、社會、技術或法律上的主張。又，大數據資料的迷思，許多人認為資料量大必然會帶來真實、客觀與準確度，這樣的迷思。

（四）、 大數據生物醫學中的倫理問題

大數據本身的發展具有未來式，如將網路上找到的資料與個人資料結合，如醫生看診能直接透過 Facebook 知道日常作息。關於大數據的倫理考量包括：

1. 知情同意：問題可能是我們管制的嚴謹程度為何，是否每步都需要先前的知情同意，還是需要一個能夠全方面概括的彈性授權。
2. privacy：匿名化應至何種程度，資料需保護到何種程度？資訊的自由與隱私的保障基本上有些 trade-off 的關係。再者，自己的資料在一群資料中時很難辨認出自己的資料在哪，如果可以被認定出來，將變成公開狀態。
3. ownership：資料的歸屬權為何，資料如要再重新放置或是分散、調整、刪除，需要一個自然人主張資料之歸屬嗎？根據資訊本身發展出的專利、智慧財產權是否有可獲利的機會。
4. Epistemology：大數據關於認識論的問題，其實如何分類本身就是認識論的問題，分類本身會影響人的資料如何呈現。所以所謂資料本身的客觀性與分類有關，分類的假設是無法完全超出社會、文化的想像。
5. Big Data divide：此牽涉到，許多人無法擁有大數據、無法進階大數據所產生的科技，這都會產生不公平、倫理的議題。

二、人工智慧（Artificial Intelligence, AI）、大數據、健康照護與法律管制

首先，法律面對的是民事、行政、刑事。AI 與演算法是大家最常提到的，至於 AI 與大數據，AI 會強大是因為數據夠大，所以這兩者是密不可分的。慢慢的我們會發覺 AI 與大數據會開始與物聯網有關係，因為物聯網可以提供非常的的資訊。另外是關於交易的保障，區塊鏈技術是最近有被提到的。下面會叫粗淺的提到 COVID-19 運用到這些技術本身，可能會形成如何的組合。

（一）、 人工智慧之法律地位：以民事法律為例

就管制的問題，我們應該先探討到底我們會將 AI 當作是什麼？進而再決定如何管制，兩者間應具有辯證關係。如果講到電子人大部分專家會認為是科幻小說的層次，或是未來也無法預見，故大部分專家會認為是高級工具。至於有無好到像奴隸般，純粹聽命於人，雖說有高端執行功能或是如青少年般會學習、成長，這個階段或許可能性，基本上以下的討論著重在工具或是高級工具的層次探討。在此情形下，AI 與人如果是夥伴關係，這是種隱喻式的說法，從法律觀點而言，比較像是利用關係。

如果說到民事責任會有兩大重點，第一個是從契約應如何看待，第二是從侵權行為法來判斷民事責任，接下來的論述會以侵權責任為主，在契約部分會較少。

（二）、　人工智慧於醫療場域之運用：新一波標準化？

事實上 AI 運用於醫療場域是可以節省人力、提升效率，像是醫學影像辨識，專家可以從監督的角度讓 AI 進行學習醫生標註方式、會做何種解釋及診斷。當然 AI 所學習的資料十分多元，可以有不同形態，所以越高端可以使用的資訊越複雜，就可以同時考量很多資訊做出診斷。甚至於醫院的醫療行政為了個人化、標準化，甚至要求最佳效率的產生，譬如將來台大醫院的 ICU、病房配置、人力配置，像這樣的分派本身，這也可以考慮 AI 是否可進入。

從大數據，數據先做處理、演算法做訓練，可以形成一種輔助系統，這是希望降低人為失誤，然而機器失誤與人為失誤得等同看待嗎？再來是希望協助檢驗、診斷、優化工作流程、降低成本、提升效率，在資料的部分，如隱私保護、知情同意、退出權這樣的考量，有偏差與歧視時應如何處理？於演算法部分，應如何透明化，如何解釋黑盒子中的運作，如何建立可追溯性，如何進行問責，到底從何開始管制，要如何管制，這是我們需要思考的問題。如台灣醫療器材管理法，於 2020 年 1 月 15 日公布尚未施行，我們可以想像，如 AI 是種軟體想像的話，就是屬於應用的軟體。或許 AI 可以算入醫療器材中，惟第三條中「作用於人體」我無法理解，如 AI 只是判斷影像，應只有影響解讀、判斷最後做出結論，很難想像以何種作用方式可以作用於人體。如果作用是很抽象的作用，因為最後 AI 的結論影響到醫生的判斷，所以醫師做決定後治療，這樣算是一種很間接的作用。如果一定要直接作用，像是某些器材裝入人體中，這樣才叫做「作用」的話，軟體的作用需要很多想像空間，這是目前我仍覺困惑之處。所以 AI 無法解釋為醫療器材，如需管制應只能當作醫療技術。

（三）、　AI 輔助醫療決策之性質

如要做研發、試驗，這兩者間具有不斷互相循環的關係，AI 做出的循環性可以做很多回，因為演算法可改寫，且 AI 可以不斷自我再提升，與藥物上市相比應用 AI 較彈性。如果從技術開發管制，像是人體研究法、試驗管理法，這些在此都會產生運作。當然侵權行為在試驗階段就產生時，可能民法也會相關，資料的保護在試驗階段需要訓練，個人資料保護法也會運用到。醫療器材管理法，即如果將 AI 當作醫療器材看待，於此也會被管制。至於上市之後，

消保法、醫療器材管理法中有特殊之規定。再者，應如何定性 AI 於醫療場域中之角色，還有何人需受到管制，造成傷害何人須負責。我們可以看到醫療器材管理法第八條第一款中提到，不良的醫療器材是指經稽查貨查驗後得知，有使診斷發生錯誤，或含有毒、有害物質，至危害人體健康者。而在上市之後，民法、消保法也有相關規定，基本上會以消費者保護法優先適用。在醫療法中，有關醫療器材、技術、應用有故意過失時，或是逾越合理臨床專業裁量情況下，也可能有適用餘地。於此會有很多法律規定在此交會，應如何系統化的詮釋，即彼此間不互相矛盾，是將來需要更細節的探討。

（四）、 AI 醫療器材的損害賠償責任與因果關係

於醫療器材法中之診斷發生錯誤，是於 AI 中較有可能發生之情形，或是依標籤或說明書，做正常合理使用時容易產生危險或，危害人體健康之虞，這點於實務案例中運用我認為會碰到一些問題。再者，性能或規格與查驗登記、登錄之內容不符，就登錄之內容不符，稍後會有判決介紹，就該句應如何適用。就病患或消費者因使用受有損害時，製造、輸入業者應負賠償責任。但業者證明醫療器材之製造、包裝、貼標、滅菌、最終驗放、設計並無欠缺，或其損害非因該項欠缺所致，或於防止損害之發生已盡相當注意者，不在此限。惟於醫用 AI 的判斷界線在哪？如何判斷？誰有能力判斷？這都是需要思考的事。

（五）、 他山之石

1. 歐盟：醫療器材法規

歐盟有 Medical Devices Regulations，原於今年 5 月要實施，但因為 COVID-19 的關係而延後。基本上該規定有把軟體放進，定義就沒有規定需要作用在人體，只要是可以被單獨或是合併使用，主要是為了醫療目的。所以我認為這個定義比台灣醫療器材管理法的定義較好，至少於 AI 在演算法的情況下，較能夠適用。

2. 美國：21 世紀治癒法＋聯邦食品藥妝法

美國管制的範圍更廣闊，包括軟體可以當作設備。第一個是，如軟體本身是作為搜集、處理、分析或是詮釋醫療資訊，像是在醫療端進行使用，不管是研究或是臨床運用。第二個是，如果 FDA 認為使用軟體與健康結果間有合理相關性，且是產生可能有嚴重危害結果的可能性，就是 FDA 的管制範圍。即如果 FDA 認為 AI 的使用雖非處理健康議題，但該 AI 的使用會對健康產生重大危害，就需要由 FDA 管制。所以手機本身如與健康有關，即有可能是 FDA 的管制範圍，尤其是從健康監督的資訊使用、搜集資料都用應用程式安裝在手機中，故手機中的應用與軟、硬體的操作會變成決定醫療照護要如何進行的決策循環，或許會擴張傳統 FDA 對於醫療器材的管制範疇。

如果用來診斷、減輕預防再都範圍內

3. 歐盟機器人民事規則

委員會希望透過此立法，但 European Commission 目前未有動作，未得到 European Commission 的支持。所謂歐盟機器人民事規則有個特點是強調 top-

down 的行政管制，希望從中央至地方運作。其中有特別提到，當機器人或 AI 的行為造成第三人損害時，由於機器人本身無法承擔責任的現況，其行為與過失都將追溯到特定的人類主體，如製造者、操作者、所有人或使用者，或者能夠遇見或防止損害行為發生的主體，須負嚴格責任，在台灣較能比擬的是無過失責任的概念。

（六）、　碰到民事及行政管制之難題

AI 的風險等級應如何觀察，台灣於 104 年 10 月時衛福部有公佈醫用軟體分級參考指引，較像是行政指導的性質。醫用軟體可以想像搜集、儲存、分析、顯示、轉換人體健康狀態、醫療相關紀錄等的處理的軟體，當然要考量是否為醫用。這些可以想像何謂醫療，如果要嚴格定義醫療並不容易，不過與人體生命健康產生危害都要包括在內，這叫做醫用軟體。在這樣情形下，該指引有舉出一些例子，電腦輔助診斷為醫用軟體，須視風險程度管制，如何判斷？從第一級為風險程度最低，第二級較高，第三級為最高，故醫用軟體的風險等級是與獨立性是平行看待，如由醫療人員做最後決策，此為二級。如果宣稱可代替專業醫療人員做決策，直接進行診斷治療功能，此為第三級，管制最嚴格，需要提供的資料也須最豐富。

1. 難題

故現在遇到的難題第一個是，上市前管制或上市後管制之比重應如何分配？有個學者認為，對於基因診斷的 AI 軟體，技術進化太快難以管制。目前在台灣用 AI 做基因診斷基本上是專家自律，算是上市前管制嗎？基本上沒未販售，如果要上市，會牽涉到 AI 本身會有演化、學習、進展等功能，有可能上市後管制會比上市前管制做得更好。但如果管制太多，廠商可能會認為成本太高而不想發展，故仍有許多需思考之部分。再者，管制醫療器材之模式是否適於管制醫療用 AI?於精準醫療之情況只會針對一個人，要如何比較？到底醫療使用 AI 是作為醫師所運用之醫療技術或醫療器材來進行管制？有些醫師可能會認為所有事情還是由自己決策而不使之上市，上市後可能會變成某種醫療器材。又，醫療過失傷害責任之表準為何？商品製造人無過失責任之標準又為何？第四個難題是，醫療用 AI 作為智慧財產權或商業機密之標的？用演算法可能很難解釋為智慧財產權，但可解釋為商業機密保護，但如提到此，又會牽涉到透明性、可追溯性、可解釋性、可問責性等，彼此間似乎有需要做 trade-off 的情形。第五個是隱私保障與公義之權衡。

2. 司法實務爭議：妾身未明的評估軟體－台中高分院 104 年醫上易字第 2 號民事判決

該案事實為有一孕婦做產檢，貝克曼公司經台灣之經銷商與診所合作，該合作內容為醫生抽血用軟體分析生下唐氏症的風險為何，當年有簽下知情同意書，即檢驗本身有不確定性。醫師就該軟體分析所做出的風險預測，認為風險不高，故繼續進行產前評估照顧，最後生下一個唐氏症、心臟血管有問題且左下肢以下無的小孩，有時超音波會擋住有可能是看不見的，這時最後是誰須負

責？這件事於 98 年間發生，進入訴訟後經過一段時間，有找鑑定員鑑定將這些數據重新解讀，也用使用該數據用軟體在操作一次，發現數年後鑑定發現結果不一樣，風險變高。雖未超過警戒值，為何資訊會不同，因為 98 年時風險是更低的，公司解釋這幾年擴充新資料，所以目前風險直提到較高的程度，當年的判讀也不應認為是錯誤的，因為資料有限，故兩次結果雖不同，但兩次結果都是對的。

於此法院見解認為，孕婦與貝克曼、訊聯公司沒有契約關係，基本上屬於消費關係。98 年時軟體屬性不明，因為當年未管制醫用軟體，也不知如何定位。後來才有醫用軟體的登記存在時，貝克曼公司才有去做登記，所以法院不認為未做醫用軟體的登記與軟體安全性有關係。98 年前已輸入使用，醫院也廣為使用行之有年，後來也得到衛福部的醫療器材許可證，顯見有合理期待安全性，最後由醫師方和解。

3. 醫療用 AI 民事責任之想像

第一個是是否能將 AI 擬制為法律人格，可以作為醫療傷害訴訟之被告主體。本身有基金存在，故當進入訴訟可以以基金作為損害賠償。如醫療 AI 有所有人或是診所、醫院使用，可能會有管理責任或是替代責任的可能性。第二個是以醫療機構與醫療人之過失為主，可以當作醫療技術想像，有學者認為醫師或是醫療院所應有檢視演算法的責任，檢視後認為可以使用才能用，無論最後結果為何，醫院與診所需通報與傳統醫療過失責任很像。第三個是以消費者保護之責任制度，可以以產品責任之觀點分析，如果是法院的態度將來不會改變的情形下，會產生何種作用？因為只要通過 FDA 之審查，基本上就會被認為符合當時科技專業水準、合理期待安全性。最後一種想像是，創造一種共同企業責任，該企業是一種隱喻的解釋，病人即使用者可能會從中獲利或受到傷害，醫療院所可以從中獲利、器材商也可從中獲利，醫事人員也有可能。是不是有可能大家一起投入資金至風險基金池，分散 AI 運用之風險，將來事故發生，可從中播現金支付。這與藥害救濟有些不同，因並非僅器材商須出錢，可能也因此反對聲音許多，實際可行性如何有待考量。

4. 大數據法律管制之測試點

於最高行政法院 106 年 判字第 54 號判決，法院見解認為，如果對資料處理的匿名化技術可以做得夠好，此時不允許退出。因資料的完整性十分重要，做研究所產生公益也十分重大，個人的利益國家不認為會是足夠大的利益，如讓某些人退出可能會形成破窗效應。

我們應如何設計？至少我們可以 Opt-out，但是法院擔心會有很多人退出，不果實際上是否真的如此，這是需要研究才能得知的。我們 強調尊重自由可以表達會較適宜，如果要保障更周全，需有選項是否給予。在一個民主社會中，我們期待政府管治、透明性越高越好，所以在大數據要強調這點，公眾才會有信任感，經過一段時間需回頭審視，到底做的決策本身是否要做調整。現在政府機關對於資訊的運用十分懼怕，對於很多實用資訊研究本身較不利。

三、人工智慧與大數據技術於 COVID-19 期間健康照顧之運用

（一）、 COVID-19 大流行

台灣經過 SARS 後政府有疫情指揮中心的規劃，在這次疫情下顯有效果。在這樣情形下，為何有部分還是無法做得那麼好，COVID-19 直至昨天全球染病人數達到 1,310 萬人，死亡人數 57.3 萬人，台灣染病人數 451 人，死亡人數 7 人。COVID-19 在全球大流行之因應，其實要做的事情十分多，於此我僅提出一些。第一是確認病毒株，開發快速篩檢與診斷模式。基本上是以科技與病毒賽跑，病毒散佈速度會多快，其實與人的行為有關，也要做各種行為的管制措施。第二個是開發特效藥物與疫苗，目前進度是大落後的，進入冬天後可能會有第二波疫情，有人樂觀想像第一批疫苗會在 2020 年底問世，但誰會想要先去注射？藥物疫苗在上市後會在幾年內下市，因為最早使用的那批會出現一些無法預料之副作用。第三是強化國家醫療系統，很多時候仍須由人面對面為細緻處理。第四是物資之適切分配也十分重要，在這些未出來的是會影響後面的規劃，至少在有效措施出來前，行為管制是最主要。

（二）、 AI 及大數據技術於 COVID-19 健康照護之運用

1. 透過健康與社群媒體等電子資料之大數據分析，包含物聯網：監測、預測及分類 COVID-19 之群聚感染
2. 利用 App 進行電子資料串連，監測個人之旅遊史、接觸史與標定疫病熱點
3. 以 AI 及大數據技術搜尋對於 COVID-19 具有療效潛能的舊藥（repurposing），或開發能夠針對 COVID-19 結構或機制產生作用之新分子：需先標定有潛能的藥，所謂有潛能的藥是以既有舊藥的資料讓 AI 操作，可能會有希望。
4. 利用 App 搜尋整理 COVID-19 診療相關之大量科學文獻
5. 運用病人特質、病例、影響等電子資料進行快速篩選、診斷、提供治療及預後判斷之建議
6. 結合遠距醫療系統進行病人篩檢、初步診療（降低醫院之負荷、減少面對面之接觸）：如利用 Chatbot 進行 COVID-19 之篩檢與初步診療，進行情緒支持、精神疾病篩檢診斷或自殺風險評估等
7. 利用 AI 及大數據技術（包括體溫感測資料、臉部辨識資料、行動通訊資料、交易資料、醫療資料等）進行接觸追蹤、物理劃裡、隔離、封鎖等社會控制
8. 利用網路資訊、社群媒體、平台進行公共衛生教育、溝通或影響
9. 利用 APP 或需快練技術進行生活與醫療資源之取得，保持對外聯繫管道，降低隔離與保持物理距離帶來之不便
10. 進行物資流動、分配之調整
11. 以 AI 對抗假消息、降低不適當恐懼或極端反應

（三）、 以 AI 及大數據技術進行 COVID-19 防治之限制

1. 良好的資料仍然太少，不足以良好的訓練 AI：目前而言，像是 google footrace 後來會有很多問題，是因為過度推估未來會產生疫情，因為使用的資訊經常是社交媒體平台的資訊，會產生系統性的偏見。這會顯示，傳統好資料的問題，在 AI 大世代很多學者都認為還是重要的。好資料不夠多，應強化資料的開放性，分享、擴展跨團隊合作，這牽涉到資料是否可共享、隱私保障等，都要去思考。現在很多大數據資料的蒐集、整理與訓練，都是為了將來，現在 COVID-19 能即時享用需要很多努力。許多監測性、診斷性、治療性及預測 AI 之機能，皆仍需長時間開發。所謂的可用本身，其實只是可以嘗試，還要做很多測試。
2. 不良或不精確的資料可能太多：大數據的雜訊與 outlier 之觀察太多，造成演算法失算，如 Google Flu Trends 就是這樣的問題。另外大量的論文仍需要篩選，及資料仍然需要依據人類常識進行選擇以及演算法之調整。
3. 目前較有實際運用可能者，仍為社會控制與資料分配之事項：隱私權、人權及正義等考量之爭議仍大。

（四）、 COVID-19 全球大流行之行為管制模式

1. 隔離、檢疫
2. 封鎖
3. 保持物理距離
4. 戴口罩
5. 避免碰觸臉部的 T-Zone（額頭、眉部、下巴）
6. 良好洗手、消毒、使用衛生紙並良好棄置等

四、在疫情期間，調整個人行動方式

（一）、 以不同程度之人際壓力改變人的行為

如何分析這些有益健康的行為，有些人認為對於改變行為方式應評估，是否值得使用，第一接受度如何，實際可遭受性如何，有效性如何，財務及負擔如何，是否有外溢的效果，還有公平性是需要考量的。故須考量，到底有無物理上與心理上的能力，物理上或社會環境上的機會，能夠形成動機。該動機可以是不加思索的動機，也可是反思後的動機，最後產生行為，這是 COM-B 行為診斷之循環。但影響人的行為有很多種方式，說之以理動之以情、做環境的調整，甚至使用 Nudging 的方式改變人的選擇結果，讓這個人在沒有太大負擔的情況下，欣然決定要做某些事。在有使用強烈誘因的情形，在目前使用乘人之急迫、輕率或無經驗、欺騙、脅迫、使用物理力量進行強制，大部分人會認為應給予十足的理由。有些人會認為可以建立新的社會規範，提供仿效的對象，經過同意之後之增能措施，這要如何區分的感受會不同。

（二）、 幾種 Motivation Nudge 的技術

第一個是告訴實情，經濟理性的當事人就會決定要做對的事。第二是做對的事對當事人而言是較容易的，如給予警告。再來是增加做好事的便利度，第三資訊或呈現方式與個人特別有關，個人的自我感被呈現出當事人會較願意

做。第四是 Framing，強調好處大家會較傾向選擇，另外提供資訊的時間，讓人能夠接收。第五是資訊本身的鮮明的程度越大越好，第六，有些時候不僅給金錢回饋，會給其他的美德回饋，會使人傾向這樣做。第七是強迫做選擇，如美國器官捐贈與汽車駕照，在考取駕照時就必須決定未來是否器官捐贈。第八是 Pre-commitment strategies，即我先與你簽約，如果能達到預期的作為結果，就可以得到獎金，否則即沒收金錢，因是自願的，就無法說是強迫。

五、人工智慧與 Big nudging 之倫理、法律與民主之考量

Big Nudging 就是透過大數據結合 AI 做 Nudging 的技術，Nudging 的定義是改變選擇結構，基本上是不應過度限制選擇，是給予不種不同推力的設計，故不會讓人覺得被強制，但會欣然使人選擇設計 nudge 的人希望選的選項。如機場中小便斗中會畫隻蒼蠅，希望男士瞄準避免造成潑濺。如果要用 AI Nudging 會如何運作，使用 AI 搜集個資或是喜好的資料，再來根據特質設計 Nudging，會使人叫願意去做。我們會發現有些人會喜歡用 Siri，Siri 給的資料是否為客觀或是置入性行銷，自己是無從得知的。如果用 AI 影響健康行為應如何操作，這樣的情況下，有人會擔憂如何整個社會都是這樣規劃，會使自主存在低，這是否是透明度太低，是否欠缺民主管制。如於旅遊網站訂了一個旅館，在下次上同個旅遊網站時，該旅館就會自動出現在右上方，故在這樣的設計下他挑的網站資訊會反覆出現，資料會越來越極端化，這種篩選資訊的效果會越來越集中在使用者喜歡的資訊上，故會越來越極端，不會有其他資訊進入。故要如何設計資訊選擇本身，事實上就是一種 COM-B。

有人會反對 Nudging，不可以用 Nudge 方式影響人生目的，從目的方向出發的 Nudge，會幫助選對，或是設計出最尊重使用者的選項。有些學者會認為讓 nudge 有合乎倫理要求，只是在方法上調整，讓使用者想有可能有別的選擇，有些人會認為是操縱。所謂強調 Nudge 影響人做決策的方法本身，什麼樣情形下有所謂可接受性存在，最擔心的是暗中做 Nudge 但無法得知。有些學者討論如果要做倫理尚可接受的 Nudge，將來是否有將某些東西轉化成法條管制，還是於詮釋法條時根據倫理解釋，這是另外一個操作方式使用類型化方式處理。我們強調是影響人決策的方法或過程，不會影響人生目的，盡量選項中有些符合他的期待，調整選項結構，幫助設計人生選項的架構，當然是符合使用者期待。另外是從民主觀點進行正當化，Nudge 本身不能太強，如果太強會很難阻擋。又，當好處越大，譬如對健康效果幫助越大，Nudge 的正當性就會越大。如果 Nudge 更可能產生健康保護效果，會認為使用 Nudge 較合適。又，從提升理性能力而言，當發現某些思考上的習慣，如非理性思維的存在，而 Nudge 可以幫助對抗非理性思維，大部分的人會是接受的。如果長期理性決策結果較常出現 Nudge，會認為較符合經濟理性的，因為長遠才是重要的，短線的常被認為是不重要的，所以 Nudge 幫助實現長遠自主是值得的。再者，最好在啟動 Nudge 時，啟動的人與被影響者要有信賴關係。如是 AI 結合，學者特別強調，資訊來源不能是中央調控，資訊最好是分散式的來源，這時資訊自我

決定權，參與也需有自我決定權，資訊的管制盡量透明，這時較容易產生信任。資訊本身須去除扭曲或污染，特徵如何設定，要讓資訊擁有者有控制可能性。又盡量讓資訊與經濟社會文化的多樣性，才不會極端。要讓資訊提供者及接受者本身有互動性、合作可能。甚至可以使用 AI 幫助資訊使用者協助分析、篩選，讓大家集體合作產生集體智慧，在所謂的資訊、電子時代，共同努力、負責，提升數位能力。

六、CODA：在疫情過後

這次疫情加速世界的數位化，強化人們對於 AI 與相關資訊的接受度。疫情過去是否我們會習慣於政府與大公司對於資料隱私之侵蝕，習慣政府與大公司善意的 Nudging 後，甚至習慣於政府與大公司更加強烈的影響措施。故在民主法治時代，如何以法律規範避免人民習慣於政府與大公司對於人民之不當 Nudging，這是將來需要思考的。從疫情時代，我們為了疫情可以忍受很多事，惟很多時候遭遇 Nudging 技術操作，可能也覺得沒關係，會認為防堵疫情較重要，惟在一晴過後制度技術演化會越來越高級，還是需要思考，再好的東西都可能有副作用。

壹、問題討論

何之行：

關於 AI 責任的問題，剛剛老師提到唐氏症案件，最後是透過鑑定，因為當時技術而後數據已經重建了，故製造商並無過失。如果將ＡＩ當成工具，即便製造商有過失，醫師本來就有專業判斷的責任，只是看製造商有無過失，其實無法免除醫師的責任。在該案例中，醫師的部分應如何處理？

吳建昌：

該案後來醫師選擇和解，醫師較無把握會繼續勝訴，不管是人的判斷或是機器判斷都有不準確度存在，如果這種不準確度是同儕在同樣情況下也會做同要的判斷，我們可能會認為並無違反注意標準，因為醫師對無上訴沒把握，故其和解。如果將來有獨立性高的 AI 出現，我有詢問過有在研究 AI 的醫療影像專家，他們會認為 AI 一定比我厲害。故我在看精神科的門診，如果有一病人詢問我，上個月做的電腦斷層檢查結果如何，我一定會將檔案打開，如果該檔案的量非常大，第一次解讀一定是由 AI 解讀，因為 AI 肯定比人準確。如果將來 AI 可以好到比其他一般專家還要準確，可以超過 AI 的可能只有幾個人，這時候可能醫療機構均視 AI 的判讀為主。所以 AI 表現既然都比人好，那有可能不使用 AI 判讀嗎？我想很少人會這麼做？如果如此，醫療用 AI 也無需負責，

因為對於當時已經是最好的，AI 本身即為自己的標竿，如出事故也無需負責，因已無法避免。

何之行：

可能會略傳統醫病關係，及傳統醫師須負的責任，會將之完全導向只看產品責任，是否會使醫師原本的判斷空間被壓縮，也可能會影響到醫學倫理。

吳建昌：

當數據非常豐富時，去詮釋或是運用、解讀的機制變的很豐富時，醫病關係可能因為中間介斷變多，會使醫病關係遙遠。最近影像醫學科有些人會擔心會被 AI 取代這些專業人士的判斷，他們將來會沒工作。

李建良：

AI 判斷是由人判斷，或是僅由電腦訴說情況，這兩者間是否有差別。將來是否會有 AI 取代人力的問題，有個關鍵性的問題是，如果能由病患自己在家打開電腦即可判讀，還是即便是醫師由 AI 分析，但是由醫師講出的話是否會有不同，特別是在心理治療中。

吳建昌：

如果是好消息的差別不大，不過如果是壞消息，除非 AI 已經精進到，講完壞消息之後會出現 chatbot，如果需要人為的互動，機器本身的感受就不同。以前在研究心理學，如小猴子在發展過程中自己長大，和有真正母猴子照顧，和絨毛猴子能夠給予擁抱，結果是有真正母猴照顧的會發展較好，其次是絨毛猴子，沒有任何形象的照顧，小孩子長大就容易有行為偏差的問題。所謂人與人的互動，在長大後，一般人碰到對人很大衝擊的資訊，自然人對於同類情感的支持，還是有差異。

陳弘儒：

有關 Nudge 的部分，在投影片中針對此有分層次，最後類別是較強制性的，該部分是屬於推力的部分，還是為非，一定要以輕微方式。

吳建昌：

再強調 Nudge 只能在改變選項的結構，可是選項都還在，對人的力道不能太強勁，才不會有違反意願的批評。

陳弘儒：

如果政府透過上面選項改變人的行為或是選項結構，手段是 Nudge，改變行爲是結果，採取這個手段改變這個行為，有時會有些副作用，該些副作用是否會影響改變結果的關係。譬如我們希望小便斗乾淨、畫蒼蠅，可能有個國家的人民完全不在意，政府再增加誘因，讓上面有感應，一直瞄準就會有分數，因為此設計後，會造成很多人排隊使用該小便斗，產生的結果可能是不堪負荷等，人民要使用而無法使用。在透過 Nudge 當作手段已達到人類行為的可預結果，是否會有結構或是想像的圖示，可以讓我們評估 Nudge 的手段是決心是多高。

吳建昌：

這是一個重要的議題，因為 Nudge 的評斷基本上是，他會把 Nudge 本身的手段會當成分析批評標的，至於 Nudge 做出的結果，本身是也是必須要被批評的，每個技術當有好壞處。回到剛剛所提到的累積分數這件事，有些人會認為已經脫離 Nudge 範疇，已經到達刺激，或許會比一般講 Nudge 還要再強一些，於此這種組合可能會有人討論不是單純的 nudge，使用這樣的科技累積分數，如果經濟實用當然這種廁所越多越好，或是規劃一天只能加分三次等，這都是分配正義終要考量的。

楊岳平：

首先，針對醫療用 AI 的部分，去年 12 月底時歐盟執委員有針對 AI 責任的報告，當中有做深入的分析，不只是針對醫療本身，尚包括一般性的人工智慧的法律責任，尤其是侵權責任。他提出幾個觀點，第一個是不建議使用擬制法人這個概念，他認為這個概念不成熟、太負責，而且有許多不可預期的後果。第二個是回歸一般消費者保護法、一般的侵權法則處理相關問題，特別提到幾個重點。第一個重點是在適用消保法時，當時科技水準這個部分，醫用 AI 會有演化的問題，故也不可仰賴當時提供產品進入市場時，拿到相關認證就表示已經符合該要件，歐盟認識到有 AI 有發展的特性，所以他們認為不能如此這樣看待，必須將時間點拉長，補充方式是他們要求包括 AI 產品的銷售者應負事後的監督責任，當然當時科技水準合理期待還是要符合，惟後還是會有持續的監督責任，否則須負類似消保法的無過失責任，對於這樣的處理方式運用在醫療用 AI 的評價為何？同時在此基礎上，他們也繼續解釋，科技水準應如何

認定，這可能與醫用 AI 的發展程度不一樣，想像夠多的可能是這是嶄新 AI，沒有同類的可相比，所以不知如何認定科技水準，一開始是使用擬人化概念。如果醫師有達到該水準，AI 在履行與醫師相同的服務時，就要達到該醫師的服務水準，這是他們所定的科技水準。惟如產品已經同時有好幾個可競爭的產品，已經變成相對較成熟的產業，那就應依照該產業之標準。換句話說，一開始以人為類比對象，之後是以產品標準作為類比對象，這運用在醫用 AI 上是否為一適當之處理方式？還是是過嚴的，我當時看完整個歐體的報告，個人認為較嚴格，相當保護消費者，但相對對產業會產生一定程度的影響。

第二部分提到Nudge的問題，我理解Nudge並非金科玉律，現實上用Nudge的法方式需經過實驗的，至少在英國於金融消費者領域是如此。而於此 COVID-19 是個疫情，在這十分緊急的狀態下，我們要應用 Nudge，但是沒有太好的實驗機會，我們只有四百多個確診案例，能實驗否都是個問題，在沒有經過充分實驗的強況下，是否會影響到推動 Nudge 政策的正當性？

吳建昌：

有些問題沒辦法馬上給細緻的答覆，基本上目前在醫療端，沒有人認為擬制法人是現在可行的方式，目前沒此設計，如要規劃一個未來才用得到的規定，可能到較接近時再規劃會較合適。之前我演講時有提到，關於照顧失智老人 AI 的發展，1990 年日本的動畫是我們現在想要達到的程度，一個照顧老人的 AI 可以將所有資料上傳到中心，連結到其他照顧的單位，如果發生問題，該些單位就可拜訪得知問題所在。這個 AI 既使現在都不一定做得到，可是當時就有想到。所以我們可能會有一、二十年可以慢慢去想這個問題，我同意目前至少在實定法規還不需要如此想相，惟學術探討是值得的。

消費者保護法當時科技水準的部分，謝謝岳平的提供，這樣處理方式對於保護消費者本身不能以當年出廠的規格，這種與時俱進的想法從法律政策目的，消費者保護法如果真的是為了保護，我們還是希望產品還是跟著別人一起變好，比較對象是決定責任最困難的事。在醫療傷害的案件中，可比較的就十分多，如有無遵循醫療慣行、個案風險利益的評估，根據法院所想像的合理、謹慎、小心的醫師，還是要看有無實際數據做行政調查，在這類的案例中，百分之八十的醫師會 產生 A，那 A 大家可能會認為是慣行。最後須統計是如牽涉到反應速度，到急診後應受到怎樣檢查或是藥物治療，離開合理可接受的時間太久，就算是不合理的拖延，有無這種資料可參考。光是傳統的醫療傷害，注意義務為何種程度，甚至精神科還有所謂可尊重的少數說，譬如憂鬱症治療是一定要開抗憂鬱劑或是已很緻密的精神分析，這可否認定做精神分析的醫師有過失？從傳統的觀點想到這些。在這種情形下，要如何比較當時科技水準，這是其中兩種，可能還有其他可開發的。至於要與醫生一樣好的話，是要何種醫師？如果是一般影像科醫師，是要與一般的影像科醫師做對比，還是要與最

好的影像科醫師做對比，最好是誰，還可能要經過海選。過程中有很多操作面上的東西要克服。如以一般作為基礎，是較有可能的，但該一般如果要以數據做基礎，以台灣的法院或是世界各國的法院單為一個案例要做這樣的搜集不容易，後來可能是想像操作較多。

Nudge 本身，如果希望 Nudge 本身要產生某種效果，而決定是否使用，從實驗要到達實定法的規範設計或是到政策的擬定，這是一種轉譯的過程，這是否為一線性化的過程，許多人抱持懷疑態度。有些人認為翻譯到後來都會變形，從公共治理的觀點，政治分析的議題，實際上是否都是有證據後才進行法律管制，其實大部分都沒有，既有的並不會分析，譬如刑法的存在是否真的有達到特別預防或是一般預防的效果，沒人會去做此實驗。很多時候只能就新的東西實驗，實驗證據本身會有限制，對於醫師而言做藥物組合，從過去的經驗這些藥物組合會有這樣得效果，而認為可能是有效的，很多時候是用此方式進行。短期做很多實驗的機會不大，那如何知道是有效的，大概預估不會有太大副作用，這就可知道大概有效，有點像臨床操作。

邱文聰：

一，回到 2009 年的案例，後來在 2018 年醫療法修正後將醫療人員的醫療責任減輕，2009 年時一審法院到底是以何種標準，使該醫生的責任被認為不存在，是因為過失部分無預見可能性嗎。

二，牽涉到 AI 與 Nudge，如果說 AI 是技術，在 AI 上應用 nudge，可以展現技術具有政治性這樣的特性。你的結論提到當疫情過後，大家會省思過去在疫情期間接受各樣的 Nudging 放棄隱私會變成例外狀態的常態化，我們真正擔心 Nudge 的地方到底為何，後來提到的怎樣在民主的框架中運用 Nudge 那些條件，但似乎都無法真正回應 Nudging 最大的憂慮，就是 program 適不適合。使用 Nudging 後，會使社會的大多數人往某單一方向做選擇，在民主社會中會造成的結果是多樣性變少、歧異性變少，也預設某種選擇是最佳選擇，同時事實上減少其他的選擇存在的機會。這民主社會中，姑且不論是好壞，存在本身極可能創造往不同方向發展的可能性，如何在疫情過後抵抗這麼強大的技術，如果真的運用 Nudging 在其中的話。有種可能性是在 AI 中故意放 bug，讓大家知道這就是會錯的，Nudging 的同時也知道會發生錯誤，讓錯誤本身被強加在系統中，可能會有人診斷時被診斷錯誤，如係強迫的錯誤造成，我們就必須面對這種危害，就不會百分之百的信賴。規範上放 bug 必然是個結果，當 Nudging 這麼強大的效用，可能會讓民主社會朝向單一的結果，我不要讓該結果出現。

劉靜怡：

在民主社會中，在做網路的規範必須要在適當的地方就是要放 bug，從憲法學者的角度而言，如三權分立這個制度他就故意製造這些 bug，他的理由也是這樣。如何讓 bug 過渡到規範面中，AI 會遇到的問題也是如此，如何讓 bug 的概念過渡到規範面中。三權分立會讓每一個權力部分都無法達到最有效率的方式運作，必須要受到其他部門的制約。

王大為：

Friction 可能較為合適。

劉靜怡：

是，就是讓他們較無效率。如果要 friction，就要讓摩擦的嚴重程度應到何種地步，才是讓人嚮往的，這就是必須規範化的。

林文源：

剛剛所說的 bug，或許大家嘆論 AI 太理想化，AI 需要數據學習，用不同套的數據或是數據本身潛在出現問題時，不用放 bug，AI 本身就有許多問題。或是使用哪些醫師、醫院的資料庫操作，並不永遠是最好的醫師，是所有醫師平均的結果，就實然面而言如何去認定限制這是本來就存在的議題，不用特別放 bug。另外是 AI 放在醫療器材的範疇談論，類似的對照案例是自駕車，最近特斯拉有發生車禍事件與剛剛我們所談論的，不同廠商根據他們資料庫所寫的數據不同情境的判讀可能是不一樣的。回到責任的問題，或許在醫療器材上可以看一下自駕車怎麼做，據我所知目前還是要求駕駛人不管 AI 如何運作，雙手還是必須放在方向盤上，以最傳統的方式要求駕駛人，醫療場域的狀況我叫不清除，提供給大家參考。

吳建昌：

目前有無國際適用 AI，的確是沒有，AI 是否需註記適合在哪用，因為訓練的資料會有這種問題。故藥物是否與全世界各國都需要做藥物試驗的確是有此考量，我不認為現在看到別的國家用的很順暢，在台灣一定會有同樣情況。另外，我同意林老師，bug 現在永遠都在其中，訓練的人員也不會比 AI 還要好，除非將來 AI 會比我們更強大，集合所有人的優點。再者，醫療器材運用基本上還是醫師要負責任，大家論述還是以目前可操作性最多，以消費者保護的角度觀之，有無可能如藥害模式處理，有些學者也以這種方式視之，既然 AI 可使大家都獲得利益，是不是可以以共同保險或責任基金的方式處理。特斯拉

目前的做法是最保險的做法，醫療 AI 短期內可能也是由醫師負責任，至於要全部人出錢或是醫療系統出錢是將來可想像的。

文聰提到，關於第一審的部分，醫師在此要證明其無過失，只要提出這樣的資料，第一審的法律裁判我沒有非常仔細研讀，我現在能想像，該位醫師做的檢查數據及影像其他醫師也可能不會認為是肢體欠缺的胎兒，再加上數據也不多，基本上風險不高，整合起來其他醫師也不認為是有唐氏症，故法院接受鑑定人意見。再上訴後，可能會有新的鑑定資料出現，醫師會有顧慮，因此和解。至於，疫情過後所謂例外狀況常態化有無方式避免，除非改成 self-program，或許會認為較無關係。之前學者提出要做 program 並不是完全不行，這樣的過程至少有參與，政府需設定 program 中極端化傾向，需做節制或調整，調整大家可以接受，或許在將來任何的網站或是平台，要做 AI Nudging 的過程這些原則是必要被遵守的，一段時間就要被檢驗，可能擔心會少一些。當然這些是理想中的情況，實際操作是否可以如此順暢這點我不確定。我的疑惑是 Nudging 本身是否會有 friction 的問題，並不是那麼容易就可實現的，可能因為人的變化，在 Nudging 的過程中還是會有 friction，可能有些人會感到反感或是有批評的聲音，Nudging 的方式就必須被檢討。這種機制自然的存在其中時，如設計在內在讓我們有機會參與 program 的過程，或許還是會較安心。

蘇凱平：

想請教有關 Nudging 與 COVID-19 的關係，我們在討論 Nudge 時，我一直在想是在什麼樣的框架下討論這件事，根據 Sunstein 與 Thaler 一開始提出 Nudge 的概念確實是不可以增加經濟誘因的，不可以使用誘因改變選項，只能在原本誘因不改變的情況下移動、設計一些選項。後來有許多的討論，在什麼樣的狀況下我們只能使用 Nudge，特別考慮到 COVID-19，有人甚至認為這是戰爭狀態，當然每個國家疫情情況不同。如果是在 COVID-19 的情況，或是在某些較嚴重的國家，譬如美國，在討論時會認為只有限於 Nudge 且是不能增加經濟誘因的，或是能增加經濟誘因的方式，譬如戴口罩會給予線上購物的優惠，或是一些更嚴格的方式，現在有人在討論違反醫病防治上規定，是否可使用刑罰方式加以處罰。如果用這樣的方式防疫或抗疫，我們討論 nudge 的意義會在哪，除了公共健康會幫助，而現在是一個嚴重的疾病，我們討論 Nudge 而非討論更嚴厲的措施的原因為何，因為 Nudge 原 Sunstein 自由家父長制設想的框架下，我的疑問是在 COVID-19 的情況下是否還適用 libertarian paternalism，的架構。與此類似的是像美國的情況，最早美國人因習慣上不喜歡戴口罩，一方面早先是他們認為口罩並無預防效果，另一方面是認為不想用強制的方式強制人民戴口罩，用建議方式。惟同一時間強制工廠關閉、不必要的企業不必經營，但於戴口罩上不強迫，不均衡的情況如何理解。

吳建昌：

關於 Nudging 之所以會被提出且具有吸引力是，在眾多候選措施中，從比例原則的觀點我們會希望手段要證明得達到該目的，在這個過程中，如果能確定能達到我要的目的，有些是強制的、有些是叫不強制的，我只是調整選擇的安排方式，你就願意做選擇，該選擇好處是大於壞處的。故能夠比例原則上通過檢驗的話，Nudge 會比強制的措施更吸引人，因為是自由放任主義中的加護照顧的態度。故在 COVID-19 這段時期，如果用 Nudge 後讓需要動用到強制的對象、數目或是嚴重度降到最低，我就認為值得，基本上可降低社會整體的負擔、人權的侵害。對人管制的技術永遠是多層次同時運用，不會單一運用，只是組合上要如何使用會較適當，我相信我們對於人的影響力不要如此直接、強力，Nudge 如果能好好運用，是值得嘗試的。再者，美國人不喜愛戴口罩這件事會讓人聯想到川普，他是為指標性人物，他一直都很堅持不戴口罩，如果學習效果本身也是 Nudge，美國在戴口罩這件事上，並沒有規劃很認真規劃出各式各樣可能的 Nudge，譬如戴口罩是十分潮流的。美國對於戴口罩有根深於文化的反對，可以從文化觀點探討口罩對於該國家的意義為何，有些國家會認為戴口罩就是有病的，有瑕疵在身上，這是值得探討的問題。

江苑萃：

我有稍微閱讀剛剛提到的判決，一審時診所與醫師是勝訴的，但還是賠償 120 萬。第一審的法律見解讓我很難接受，在一審時有提到，左膝蓋以下下肢缺損普通超音波的診斷率只有 22.7%。用很老舊的超音波機器應該還是看到的腳掌，看不到腳掌的機率是 70%，這讓人很難接受。第二個是母血篩檢唐氏症，不管是較早期母血篩檢的技術，或是現在很熱門非侵入的檢驗，給的都只是一個百分比。以該案而言，篩出得到唐氏症的比率是低於 270/1 就會被認為不需要做羊膜穿刺，這個產婦本身篩檢結果是 38,000/1 機率得到唐氏症，因此醫師不推薦產婦再進行羊膜穿刺，理由是做羊膜穿刺有 1000/1 機會會流產，但做羊膜穿刺的結果就會接近百分之百正確，因為是直接取到胎兒的組織。從該判決可發現，做篩檢的公司只有這種結果滿取巧，這樣的檢驗結果已可供醫師參考，故同意上市，篩檢結果給的是個比率，合乎當初申請查驗的標準，我們就認為沒有過失。可是對醫師而言有很多選擇，可以選擇推薦羊膜穿刺，也可能推薦產婦做這些篩檢。一審之所以認為醫院與醫師無過失是因為，既然醫師有推薦做母血篩檢後，且依據篩檢建議可不做羊膜穿刺，診所醫師，醫院與醫師無過失，既然醫師有推薦，依據篩檢建議可以不用羊膜穿刺，因為篩檢數據低於 270/1 的風險，這樣的結果足以可以上市，怎麼會該標準因此變成醫療常規，可以不用再推薦產婦做羊膜穿刺，醫師可以從這個責任脫免。未來使用 AI 輔

助診斷系統，會不會也是給一個數值百分之多少是正確的，只要跑了系統是百分之多少，依據此醫師可以從此責任脫免。

吳全峰：

剛剛在討論此問題時，一個是從消費者保護法、一個是從醫療法，當 AI 介入後，產品與醫師判斷的界線開始模糊，但於模糊的過程中，在法律上的處理又將標準拉開。即消費者保護法本身走產品，到了醫療法第 82 條又往另外一套走，當 AI 本身過程造成產品標準與醫師標準中間開始模糊時，法律的處理又將此拉扯開，如何解決這樣的問題。

吳建昌：

若不是必要我們不會去冒羊膜穿刺的風險，這是 risk 與 risk 的 trade-off，需要資訊的過程中，有時在醫療上就是冒險的過程，必須在沒有資訊冒著天然風險，與有資訊冒著人為風險中取捨。出事之後，冒著天然風險的會認為醫師當時為何不冒人工風險，有可能可以及早發現，發現結果可能是做流產手術。在此本身的權衡應以何種方式看待，在醫審會中很多專家會認為在這樣的果程中，醫師要不要冒險做這樣的手術、診斷，如果大部分醫師認為毋須冒該風險，或許結果出現是令人遺憾的，惟應不會認為醫師的判斷有誤。後來會有生產事故救濟條例，基本上是為了讓婦產科醫師的困境以基金作救濟，該基金醫師不必出經費，大部分是由菸捐提供的。這部分很困難，會不會醫師只要把醫療常規轉化成只要 AI 認可，醫師即可免責，如果真的有此發展的可能性，將來應改成無過失責任，如果無法達到該情況下，應該要有可追溯性的設計與咎責。在當時會有類型化的發展，在某種類型的案例中，醫師要負責或是 AI 負責，政策上應導引至這樣的方向。關於全峰的問題，到底要如何設定標準不是那麼容易，如果用醫療技術想像 AI，基本上醫師做出的判斷就應該負責任，至於判斷是否可通過同儕審查，認為有盡到注意義務，注意義務應用哪些檢視，如其他人使用 AI 做出的判斷與你不同，我們會認為判斷有問題，還是可以用醫療過失醫療法 82 條，當然檢驗的標準比以前多，或許檢驗的標準多讓醫師豁免機會較大。

吳全峰：

說不定最後只有一兩個醫師可以根據 AI 做出自己的判斷，而其他的醫生都是依循 AI 的決定做出判斷，這時是以這一兩個做出自己判斷的醫師找出 AI 判斷錯誤為主，還是回到大多數醫師已經無法判斷 AI 是對或是錯的標準。

吳建昌：

所以是未來可能沒有任何醫師會違反 AI 的判斷，最後常規都被大家遵循，如為自己判斷則會變成違反常規。的確有可能，我認為還是需一段時間，將來如真的發展到此階段，那將來的醫療傷害本身，在台灣我會全力推動無過失責任，希望用基金的方式，既然大家都是使用 AI，我們希望 AI 繼續用，錯誤可能會發生。

王大為：

如果以後人無法判斷，但有其他的機器幫忙判斷，會是怎樣？

吳建昌：

醫療 AI 是一個測試點，因為希望 AI 幫忙決定通常是較無關緊要，但又反覆的苦工、反覆檢視由 AI 進行，人類決定的是重要具有傷害性，既然區分是重要具有傷害性，醫療大概多數人會認為如此。但那是想像、與人性發展，如醫用 AI 越來越多，用藥參考會不斷跳出，醫療影像的解讀也會先跳出，為了保護自己而為防衛性醫療。將來的醫療常態可能大家會認為看 AI 就夠了，台北榮總已經有 AI 影像的門診在運作。

李崇菱：

Sunstein 四月的時候有寫篇文章，裡面的沒有提到 Nudge，但有有講到如果川普帶口罩的會改變的他的社會意義。回到 Nudge，實證研究在食品有用 Nudge 鼓勵人民消費較健康的食物，這想法是有些時間，故他們可以參考很多國家政策是如何影響，後來發現食品有出現層次，對社會影響並非一致性的，高收入或高交易會改變行為，但在低收入或低交易則否，事實上這種差異是結構性不正義的問題，如果是使用 AI、Nudge、醫療行為，COVID-19 的案例中有很多是有慢性疾病的病人，如在高收入的國家中，有慢性疾病的人士較中低社會階級，如果搜集這些資料對未來在管制病時是好的，這裡還是會有社會不正義的問題。針對疫苗的問題，有個社論可以顯現出對政府是不信任，對科技信任，他們提到如果有疫苗出來，應優先給黑人使用，因為在社會中是相對弱勢的。

吳建昌：

使用科技 Nudge，科技本身本來就有 divide 的問題，剛剛提到就是所謂的 Nudge divide。Nudge divide 本身是好事還是壞事，也是值得論述的，有時我們會擔心 Nudge 太強大，Nudge divide 會有有些人不會被 Nudge 影響到的情況。如果政府的影響我們都相信是好的，國家應設想不同 Nudge 作用的方式，有人說就說 Nudge 的處理方式如傷害身體健康的可樂，放在最伸出，對某些人而言，將櫃位放在某些地方就會有影響，不同屬性做不同 Nudge 會稍微降低結構性不正義的問題。至於疫苗，在美國與歐洲有許多關於疫苗的爭議，有許多反對疫苗運動的團體力量也很大，很多人討厭疫苗，也有人喜歡疫苗。討厭疫苗的人會認為優先給黑人使用是將他們當白老鼠，下市的結果可能是黑人的副作用很多而不用。如是喜歡疫苗的人，會認為優先給予黑人使用是因為他們是社會弱勢者。

李建良：

一、剛剛提到在 AI 中設 bug，科技部的人工智慧科研發展指引有八點，有一點是安全性，設置 bug 在其中是否意味我們不想讓 AI 太安全的意味，又剛好與 AI 更安全的觀點是否價值上會更衝突。是否將來自駕車零事故就予不上市，還是選擇性，這是我們對安全性與科技部人工智慧科研發展指引思考，我認為有些價值衝突的部分。

二、將來要用無過失責任，用基金解決問題，當我們無法解決問題時，某種程度我們就適用集體性方法解決，由大家分擔。於哲學而言，某種程度也是自我放棄。當問題無法解決時，會不會有避難至 AI 免除責任。

三、AI 是否為一醫療器材、醫療技術，器材與技術是兩個概念，還是一體兩面？而技術物是否會是器材加技術的概念，文源老師談 AI 是從技術物方面研究，在社會學的領域中有所謂「技術物」的概念。惟於法律領域中，談到醫療器材有三部法規規範，一個是醫療器材管理法，第二個是醫療器材管理辦法，藥事法是母法，第三個是醫用軟體指引，基本上是行政指導，這只是目前規範現狀。醫療器材的要素，重要的是作用於人體，政府的措施是否皆為醫療器材？公部門的措施只要作用於人體，可能可預防疾病，按照這個定義是否皆為醫療器材？

四、Nudging 到底是一個新的名詞還是本來就存在，只是在創造名詞？這是一個現象，在辯論是否為 Nudging，還是是一直以來都有的一種做法，但這種做法會有各種現象，因為有強調後變成是一個接近的方向。最後，法律到是否應介入 Nudge 規範，確實有兩種方向。一個是強化 Nudge，另一個是移除 Nudge，我們本來就有置入性行銷的規範，只是置入性行銷規定的範圍會較窄，事實上我們也可以認為是某種程度的置入性行銷，置入性行銷可以規範的話，Nudge 也可以規範。

第二次會議紀錄	
時間	2020 年 8 月 31 日（星期一）
主題	人工智慧與社群媒體分析
講者	魏志平（國立臺灣大學資訊管理學系教授）
內容摘要	

一、研究領域

（一） 人工智慧

唸書時有在管理學院研究人工智慧相關議題，在當年並非一個熱門的題目。後以資料為導向變成人工智慧中一個大的脈絡，當時叫做資料判斷，這個詞大概於 1990 年前期到中期被界定出。故機器學習現在所談的基礎，大概就是資料探勘的基礎，做所謂的監督式學習的概念。我也做很多文字相關分析、文字探勘、資料檢索、社群媒體分析，這個在技術層面，過去 30 年我大概就環繞在類似的主題上。

（二） 應用領域：商業智慧、專利分析與探勘、醫療資訊

於應用領域，在管理學院是做與商業有關，比如說分析社群媒體資料，目的是為了解決行銷決策的問題、品牌管理相關問題，故下文會談及品牌個性如何利用社群媒體資料能夠偵測出來。那財務上，這兩年做了兩個相關研究，如一間新創公司的模型為何，我們使用許多數據在操作，包括募資、企業內部資料、社群媒體上被討論的內容等等。清華科技管理學院有許多同事從事專利分析，將資訊的技術用到專利的分析上，可以有各種不同層級，譬如專利層級。譬如企業中，美國專利每四年、八年、十二年需要維護，而到底哪些專利需要被維護，我們使用 IBM 過去關於維護、不維護的資料想辦法產生某些變數、建立模型，以產出該專利是否應維護等等。這講的是專利層級，情境是企業中智財管理的範疇。如是科技領域的層級，譬如奈米科技，先定義好奈米科技，透過某些關鍵字跟類別的搜尋，找出科技機會在哪，即哪些區塊過去沒有很多人研究。

醫療部分，分為兩大部分，一個是我們曾經使用過健保資料庫，於 2012 年台大的一個健康加值計畫與藥學系老師合作，利用健保資料庫裡的 100 萬人抽樣檔尋找藥物不良反應的訊號。並非特別著重急性藥物反應，比如吃感冒藥會頭暈、嗜睡，我們著重是長期反應，如哪些藥會導致癌症、心臟病、心血管疾病、腎毒、肝毒等。所以使用 10 年的數據串連人的資料，較能偵測到這種長期性不良反應。另一部分是做文獻探勘，使用美國醫學文獻，到目前為止是 2500 萬份以上，想辦法理解每一個文獻中的概念與關係。譬如某篇文獻中是討論 A 藥來治療 B 病，A 藥可能跟某基因有關。另一篇文獻可能探討該基因跟另個基因有抑制作用等等。我們將關係萃取出後做圖論的推論，即有無未知關係，可以從現有關係推理出來。典型的例子是「舊藥新用」，某藥當初的適應症、能解決的問題可能是某疾病，可是他可以被再提出治療其他疾病。較耳熟能詳的

例子如落健、柔沛，原是掉髮禿頭用藥，當初一個是高血壓用藥，一個是前列腺用藥，後來發現二藥的服用效果可以抑制禿頭。再者我也有做機器閱讀理解，這幾年較熱門的人工智慧題目，也就是輸入一段、一篇文章、一個問題找出答案。

二、社群媒體資料的價值

社群媒體包括社群網站、產品評論、論壇、部落格、微博等等，已經是人類常常利用分享自己日常生活發生的事情當下的心情、經驗、看法、探度等等。那社群媒體資料目前看來也成為許多領域重要研究議題的一個替代數據來源。特別在管理上面，很多過去我們可能需要用問卷去問的意見，我可能可以透過社群媒體資料的分析，得到一個檔案。

社群媒體資料已成為許多領域重要研究議題的資料來源：

（一） 消費者行為分析、行銷管理：

如於在消費者行為分析、行銷管理上，可以透過分析顧客留下來的產品評論、個人經驗也好，我們可以試圖了解消費者對某些商品服務的消費金額和情感態度，進而可以做到消費者行為理解和行銷品牌管理。

（二） 股票漲跌預測：

30 年前在管理學院念書時利用基本分析來分析這個股票到底好不好，後來就有技術分析。不管是基本分析或技術分析，技術分析大概以天為單位，也就是我分析今天的線圖，可能是使用過去的線圖預測明天股票可能的漲或跌的趨勢。現在許多人在社群媒體上，甚至一些與投資財務有關的社群，讓大家大量分享對於股市大盤趨勢的看法、對於某些產業的想法、對於某一張股票分析的結果，這些資料是以分鐘可以收集到。故這十年來有不少研究就是做專門在收集社群媒體的資料，如 Twitter 上的數據以每 15 分鐘登記一次，決定下 15 分鐘這張股票的漲跌。可以認為買股票已經是以分鐘的決策為單位，而非以過去我們看昨天的線圖來決定明天的趨勢。這大概是股票漲跌預測。

前陣子我做完一分析，即 IPO 股票新上市對於財務的人而言，他第一天的漲跌是他重要的議題，叫 underpricing，也就是所謂的發行價一般而言會低過市場的預期價格，所以看第一天漲多少，想知道到底是什原因，或有沒有預測模型可以來預測，所以我們拿財務上的指標再加上社群媒體討論 IPO 股票數據，可以看到，以這類管理或財務議題，那這幾年來利用社群媒體上面的資料將他們過去的模型往前推一步。

（三） 政治輿情分析：

政治人物常透過社群媒體的觀察網民的看法，網路上的風向、談的議題、對某政黨的看法、政治事件的看法。在資管中，某些文獻也表明，他們利用過去 2008 年的美國總統大選觀察社群媒體對於美國總統大選的影響。

（四） 藥物不良反應：

在公衛藥學領域，過去仰賴衛福部食藥署的不良反應通報，一般使用者使用某藥覺得不適，有一些你可以觀測到的不良反應，可以主動通報。藥廠如果知道有這個反應，他必須義務性且強制需通報。惟我們也分析過美國不良反應通報系統上的數據，那他們的數據的落後程度，也就是說很多人要再去用應用程式登錄，並非他所習慣的日常使用。可能較傾向在 Twitter、Facebook 上面就寫我昨天使用哪個藥，我現在頭痛的要命，所以有些研究抓 Twitter 上的數據或是 Facebook 上的數據進行不良反應早期偵測。他們也發現會比利用食藥署不良反應通報系統的數據，可能可以早上幾個禮拜或幾個月可以更快知道不良反應訊號在哪。

（五） 精神疾病偵測：

上學期與學生用 Twitter 的數據做可能憂鬱症的可測，於 Twitter 上面收集很多人的數據，抓完回來那哪些人可能是憂鬱症可能不是，怎麼辦呢? 如果我要發問卷也不知道那些人是誰，也要經過學術倫理審查等等，將曠日費時。較簡單、較沒有倫理問題，我們使用醫生看這些人寫的內容，用他的專業判斷，懷疑他是否有憂鬱症傾向，所以找了三個住院醫師，大概 coding 出來幾百個使用者，我們用了某些特殊的關鍵字比如說 depression、depressive 這幾個字，所以比率上就比隨機去抓回來的多。最後我們分析文字內容、情緒變化，建立一個預測模型來偵測精神疾病可能的訊號。

邱文聰：

如何確認 Ground Truth?

魏志平：

於藥物不良反應，當時本來是希望藥學系合作的人能幫我們看，惟沒辦法看。所以我們能做的是針對我們找出來的訊號，去Micromedics 的網站，該網站有某個研究說 A 可能導致某個不良反應。他們就分級，比如說這是一個大型的調查結果，這是一個動物實驗，故我們只要認為已經有研究在談這個訊號，被我們找到，就當成他是 Ground Truth。那至於有沒有很多東西我們抓到可能是，那可能還沒有研究，那我們希望有，也就是說 precision 是 underestimated。

至於精神疾病，我們討論過幾個方案，一個就是在我們周邊的人想辦法招募幾百個人，願意把 Facebook 或 Twitter 的帳號資料，讓我們的程式存取。同時做某些精神疾病的檢測，但這曠日廢時，需要經過研究倫理審查等等，而且這些人可能也不願意。所以我們能做的就是，我們請醫生看著他的寫作去判斷。那有沒有可能誤判，有可能。有沒有可能錯過掉，有可能。但是我們就只能在這樣的概念之下去做 Ground Truth。

邱文聰：

這樣訓練出來的演算法，他是要趨近於一個精神科醫師的能力去判斷在社群媒體上某一句話的人，他是有精神方面的疾病，是要達到這樣的目的嗎？

魏志平：

我們沒有要取代任何精神科的概念，我們在做很多偵測或預測模型，任何一個專業每天要面對這麼多訊號要處理，其實是很恐怖。如果可以幫你過濾掉某些或是確定這些訊號優先順序，希望能夠幫助專業人士減輕壓力。就像藥物不良反應偵測，我們列出我們系統中偵測到我們認為訊號最強的，也就是機率最大的。當時一個林藥所的老師，他認為我覺得這個訊號很特別，他要做藥理分析。他做完藥理分析最後認為訊號是正確的，所以就發了一篇文章，故我們沒有要取代任何專業的人。

吳舜文：

我想追問剛剛的問題，因為我認識的憂鬱症患者，在症狀嚴重的期況下，他們任何事都不想做。這種偵測是不是只能只能偵測到早期的一些可能徵狀，所以真的被診斷為憂鬱症的患者，可能根本不會使用 Twitter。

魏志平：

是的，在執行這個研究計畫的學生本身是患者，所以我大概清楚剛提的情況，當他真的在一個比較糟糕的狀況，他是沒有能力和外面溝通。當他好一點的時候他可以溝通。

吳舜文：

可是有時候，說不定有些較擅長表達的人，是抒發壓力，也許他不是真的有該傾向。

邱文聰：

就剛剛的問題，我其實是好奇，像第三種藥物不良反應的演算法，他聽起來好像並不是已經能夠去找到一些隱藏的因素。而是你們用台大的 database，看有沒有暗示，這邊找出來的不良反應的某些訊號或症狀，有沒有在社群媒體上出現。如果只是這樣，我需要有一個演算法嗎？還是，我只是用關鍵字就可以?？

魏志平：

我簡單講一下不良反應偵測當時的邏輯，我們在捕捉的是，用藥後後續看診中間，有沒有強的關聯性。這個關聯性你有很多不同的算法。我們當時想辦法把這些的指標建一個機器學習的模型，故當很多人吃了同一個藥，半年後、一年後都發生某個問題，沒有吃藥機率上明顯比剛剛來的低，顯然這個訊號是強的。我們是在這麼大的搜尋空間裡，想辦法去找出訊號比較強，然後我們做一定的排序，故希望排在前面的可能就是真的訊號。所以被我們找出來的東西有部分可能是已知的不良反應。希望有一大部分是真的都不知道的不良反應。沒有用社群媒體的資料，也不是已知吃 A 藥會得癌症，去健保資料庫搜尋，我們不是做確認，比較像是做探索。

李建良：

在 Twitter 上收集資料，有經過當事人同意？或是如何收集？

魏志平：

因為 Twitter 有開放 API，所以我只要下關鍵字，我找到某些人。那這個帳號只要他願意公開，就直接收集，故當事人不知道他的資料在我們的資料庫裡。

李建良：

第二個問題就是要做什麼？會不會有商業利用的可能性？

魏志平：

如果藥物不良反應，我覺得有，只是需視藥廠願不願意付錢。如果會讓他的藥讓他暴露在越高的風險可能不願意。曾經有人跟我說我們應該去找藥廠，但沒有人希望把自己的東西找更厲害的演算法把弱點揭露出來。但是誰可能可以用？食藥署可能可以用、做藥學研究的人可能可以利用。精神疾病方面就是純然以學生的狀態，對題目有興趣，而且顯然他是非常有病識感的人。他也覺得有沒有可能從社群媒體當中。我相信沒人要用最後一個研究的結果。那學校方面可能可以用，我不知道。

李建良：

最後會建立預測模式或模型，那這模型應該會有很多人想要用。

蔡政宏：

想確認一下，剛講到股票有一堆資料，政治輿情有一堆資料。那這些數據要建構成模型。在講自然科學或哲學，在科學哲學中也是有一堆資料，那資料是自然界的。那科學界是根據這些數據要建構模型。但在科學中，模型最主要有兩個目的一個是預測，另外一個是解釋，要去能夠說明性向的連結。像資料漲跌和政治輿情這些找出來的模型，好像與科學的科學模型不太一樣，因為至少要做兩個。像看古代太陽和地球之間的關係，古代人可以建構的是一個地動說，地球是宇宙的中心，他收集的數據是這樣，但我們現在知道不是。所以只是純粹預測，可能在沒有太大的誤差下，兩個預測都可以，但到現在我們知道解釋完全不一樣。所以我想知道從資管的角度來看，建構模型最後的目的純粹只是做預測，但是對解釋他其實是不在乎的？

魏志平：

在管理中，這兩派的人是分開的，建構解釋性模型的人目的是做政策依賴跟管理依賴。那建構預測的人，是要利用預測模型，他要加速他的某些決策，這兩個完全不一樣。所以做解釋性模型的人，在假說的推論背後的理論基礎要求十分嚴謹，也就是說從 A 推論到 B，從 B 推論到 C，所以你有一個假說，A 跟 C 的關係是呈現正向或負向。或甚至一個倒 U 的關係，要講得非常清楚。目標是當我已經知道推論結果，資料呈現，應是比關聯性多一點點的相關的可能性。所以這時我知道他們有負面關係，所以我只要操作 A 應該預期可以達到 C。預測者目標是想辦法提供準確的模型，至於變數本身，他是相關的關係還是共同因素，可能是影響 AB 共變，那 A 影響 C，所以我抓 B 以為他影響他，其實不是。

我常跟學生說，你覺得身高跟你考上理想中大學有沒有關係，可能真的有關係。如果真的是相關的，那大學聯考就改成身高測驗。可是可能不是相關的，可能是因為家境好，他願意投注在教育資源的能力比較好，同時因為家境好，所以營養比較充足。所以源頭是家庭經濟、社經背景，導致教育資源投入，與身高高，最後成績好是因為家境好不是身高。所以做預測模型，然後盡量把所有可能變數放進來，目標是建構一個真的具有一定準確度的模型。因此建構預測性模型的人比較不會去說明和在意，變數放進到底是因果還是剛講的關係，反正我人為有關係就放，模型自己會去過濾。

三、報告大綱

（一） 情感分析：個人特質

（二） 使用者輪廓分析

（三） 品牌個性偵測

四、情感分析：

（一） 情感分析(sentiment analysis)技術

1. 概念

情感分析是各位常聽到的正評負評，有很多層級，最常被提到的就是文件層級或者是概念層級的情感分析。那所謂的概念層級是什麼意思?下圖是使用者寫的評論，描述這是很棒的飛行經驗，娛樂設施雖然比較少，但是食物很好，工作人員很親切等等，這是一個最佳選擇。所以如果以整體文件的情感態度，雖然標註紅色的字不滿，但是整體是正評。所以這叫做文件層級，整體文件是正評或負評。第二個是更細微的分析，除了知道正評負評，我可以算出 60%的人對這家航空是正評，似乎對經營管理沒太大幫助，想知道他們到底在講內容為何，他們在意的是正還是負。故以這一篇評論來講，談到四個面向，所以 aspect 可以叫做概念可以叫做面向。第一個是娛樂選擇，這個人認為是負面的。座位大小是正面、食物是正面、工作人員是正面。所以各位可以想像，如果將所有 EVA AIR 的評論全部做左邊文件層級，我只能夠得到正評比率有多少。如果做右邊概念層級的分析，經過彙整之後可以知道多少人討論了座位大小，而在討論座位大小裏頭有多少比率是正評，有多少比率是負評，有多少人在講食物，有多少比率是正評，有多少比率是負評。當將兩家航空公司都做概念層級的分析之後，我可以 dimension to dimension 比較整個市場上，對於兩家航空公司在不同面向的優越勝敗到底為何，可以得到更細緻的分析。我們剛在講情感態度、正評負評，可以是 negative 或 neutral。那情感分析也可以看一個人在文本裡表現出的情感狀態，比如說他是生氣的、悲傷的、快樂的。也可以拿來做某些客觀的分析，這個評論是主觀的，這個發言是客觀的等等。所以有各種不同面向。從內容上他想要表達的情緒、態度、主觀都可以涵蓋在這個大的範疇裡。

陳弘儒：

老師您剛提到主客觀分析，他的區分是，主觀是指他的 active 嗎?

魏志平：

主觀會描述比較多個人的 active，客觀就是對一個事件的描述。所以比如說，看餐廳評論，有人講菜色不好，他每個都很具體分析，我覺得是客觀的。但是有人直接說太糟糕了，我根本就不想來，我覺得他沒具體講出任何東西，甚至講一個服務人員態度不好，如果我只講態度不好，我也認為這有點主觀。但如果講說他點餐時一句話都不問，我提問，他就跟你講說你問這麼多幹嘛，所以我覺得不好，這就非常客觀。他提供具體的事實。

2. 實務應用案例

那情感分析在實務上的應用是做輿情分析，不管是在企業界對於產品服務，在政治界對政治人物、公共政策的態度，我們都可以用剛剛所提及的概念去做。那在金融股票的應用，前面提到利用 Twitter 上網民的分享，大體上作正、負評，就可以知道市場是朝正面的態度還是負面的態度，可以用此資訊來協助提升股票的分析或漲跌預測。電商系統或社交網絡中，現在很多聊天機器人他們很想將情緒放入，也就是說，如果今天跟聊天機器人討論，已經有點不耐煩，他還跟你說我聽不太懂你在講什麼，那這個產品很糟糕。他如果可以偵測到你有點悲傷，不要再告訴你很多客觀事實、安撫情緒。如果他能夠做到，這機器人可能擬人化的程度會比較高一點。那聊天機器人除了理解人對話的語意跟如何回應之外，情緒或情感分析的技術可以協助他更擬人。

（二） 文件層級的情感分析技術：

在文件層級，大家可以想像一個文章，為什麼讓你覺得他是正評還是負評，可能跟他使用的字詞有關。以我剛提到的例子而言，藍色屬於正面，紅色屬於負面，因為紅色、藍色比率看起來差很多，最終的那句話也就是最佳選擇，讓我們就可以判斷是高度正面。所以在技術上比較常見的是把他視為文件分類的問題，也就是說，每個文件想辦法萃取重要的關鍵字，代表很多文章中比較常出現的字。再去評估每個字在文章中出現的重要性，比如說出現越多重要性越大，最後會變成一個矩陣，可以建立預測模型，太細節不多講。通常再萃取這些字的時候，做自然語言處理有一派喜歡對詞性去做分析。我認為在表達情感或情緒，一般而言可能都用動詞形容詞，最多加副詞，很少用名詞代表。舉例，你會說 I'm very happy 你不會說 My happiness is high。你不會形容東西變成一個名詞，然後再用一個等第的形容詞去描述。這是我們看文件描述情感通常用形容詞、動詞，最多再使用副詞。我們先把文章的每一句話去做詞性標註，留下我想要的留下詞性的字，接下來進行重要性評估，那留下文章代表的關鍵字或特徵字詞再往下做。

邱文聰：

所以標註是你們標註？

魏志平：

詞性標註是使用程式標。

邱文聰：

那標完之後判定正、負是由人來做？

魏志平：

對我們一定要有訓練數據訓練。

邱文聰：

這個 trainer 是你們自己?

魏志平：

對，我們自己。

邱文聰：

比如前面 EVA AIR，法律人應該比較知道有些文字稍微轉換一下，意涵可能就整個變了。你可以先把正面寫在前面，即經濟艙座位大很舒服、工作人員很親切等等，可是機上娛樂實在是太少了。

魏志平：

那可能就變負面。

邱文聰：

所以這邊也是要你們主觀判斷？

魏志平：

這邊是主觀判斷。但在管理學院做 coding，應該還算是比較小心，假設我們要對這個做判斷。實務上是，這些網站上都有心得，我們就不用判斷，也就是這個人給了三顆、四顆、五顆、兩顆、一顆星。我們把四顆、五顆跟一顆、二顆當正跟負，如果沒有的話，在管理學院，會做 coding 的訓練，我想很多做文本分析的都是這樣，我們先拿二十篇出來跟大家解釋正評、負評概念是什麼，然後請 coder 開始標，標了二十篇對答案。對完答案說第三題大家意見不一致，其他幾題差不多，我們針對這些不一致的開始討論，所以有可能像老師講的，他把負面評論寫在後面，顯然他很在意這件事情。或他的轉折是怎麼寫的，那

大家搞不好最後有共識認為他標得有道理。所以我們經過一次訓練數據，再做第二次訓練，一直等到匯集後之後，看你設定為多少， code 就直接標他是什麼。那偏技術的人就覺得這個太麻煩了，乾脆每個人找十個人標，最後取共識群就好。我們舉辦投票，五五波的全部拿掉，我就可以得到兩個極端。這是大部分人同意正評和負評，所以標註有不同做法。

邱文聰：

所以是二元？

魏志平：

對。那你也可以把它變成不同等第。

何琳潔：

想請問語跟詞跟句子邏輯上面，像上面這則比較完整，可以看出比較通順。但一般，尤其是年輕人在寫社群比較短，而且有時會用一些詞，如酸言，那機器要怎麼判斷？尤其是句子短，前後看不出來，要如何處理？

魏志平：

我們也看過一個評論這樣寫「zzzzzzzzz」，可能這產品或服務對她來講像是要睡著了，看起來是負評。以我們做文字分析，發現 formal document 的準確度一定比較高，比如說文獻探討或文獻分析。機器讀醫療文獻，複製編輯器（copy editor）幫你看過十次，應該不大會有錯。可是走到社群媒體就有很多問題。所以有一些應該被處裡的，比如說應該放 emoji 的一個字典在背後。比如說笑臉。會有一個 emotion 的 dictionary，所以就可以把那些東西就轉換成某個 token，像「nooooooo」可能會有一些規則去處理，比如說 repeated character，就把它縮到到字典裡有的字，就解決了問題。

何琳潔：

在擷取時會不會連帳號的個人資訊，而判斷他們的用詞用語。

魏志平：

應視研究的目的在哪，如果我們只做回顧，不擷取其他東西，因為太麻煩。但如果我做以人為單位，我可能希望他揭露了、公開某些資訊，我可能有興趣做。但是我們會盡量做合乎 API 就抓。如果我們做的有違法，請大家再跟我們說。

何琳潔：

老師我的意思是因為不同群體或是不同年紀的人用某個字的意思可能會不一樣。那剛剛老師說把它放到 TRAINING 或 DICTIONARY 的時候，會不會考量他們的背景，再把他放進來?

魏志平：

不會，太複雜。

（三） 範例：情感詞典，

傳統算法每一個字當成獨立，你可以想成這個文章變成二十個關鍵字。至於關鍵字的前後次序關聯，我們完全不在意。這是最傳統的模型。這個大概是深度學習之前只能做到這樣。那在深度學習之前我們偶爾會做某些事情。就是說你只要出現一個字叫「organization」，假設這個字我認為重要，那可能出現叫「firm」這個字，一個組織、一個公司其實意義差不多。那一頭出現「happy」，那另一頭出現「joyful」可能都是一樣，差不多。所以同義字或相關字在傳統模型上，視為不同的 token、不同的元件，準確度基本上是會有問題的，特別當你的文件長度短的時候。像我們如果分析學術文章動輒三頁、五頁，問題比較少，因為你常在裏頭換不同的字。可是如果是一個貼文，以 Twitter 來講，是 144 個字大概是 3、5 行，這時候你會用不同字描述不同概念的機率不高。所以文章越短，一義多字或相關詞的問題沒這麼嚴重。那到了深度學習，做這種工程的人，要去理解每個字的意義其實不容易，但如果我們可以把字投影到空間上，讓相似的字在附近，所謂相似的字是在語意上也好、文法上的特性相似的都聚在附近。我們自然就可以做運算法。比如說，所有形容快樂的都在這一區，形容悲傷的都在那一需，形容人的都在這一區。還有一個很有趣的技術是，經過深度學習去學每一個字在多維空間上的位子。他們發現其實很有趣。這邊出現東京、DC、台北、北京，這邊出現日本、美國、臺灣、中國，就是首都跟國家的關係。那他們用這個來解釋他們空間的投影，除了抓到可能的意思之外，連相對的關係其實都可能在這多維空間上。所以到後來的模型到深度學習在理解文字這件事，就不是走 Vector space model，即每個字當成一個 token 單獨處理。他就是把字投影到空間上面，那字一旦能夠投影，那在經過複雜一點的訓練，現在可以做到句子投影到空間上面去。所以「我今天很快樂」跟「我今天感到非常愉悅」這兩句話，我用的詞可能不太一樣，但是這兩句話意義差不多。所以如果在空間上，這兩句應該投影到同一個地方去。那句子一旦能夠投影之

後，文章就能投影，因為文章是由句子所組成的。那這個模型不知道大家有沒有聽過，現在最常被用到的，BERT 的模型，它是 Google 用的非常大量的 corpus 的模型，BERT 想辦法訓練文章中句子上下文的關係，將句子投影到空間上。現在很多的深度學習以此基礎，往下去做。

蔡政宏：

做文本分析是否會遇到情況是，因為聽起來在多維空間裡面可以被脈絡化，脈絡化可以被計算化。那會不會有個詞，可能同時佔據比較好或比較不好。我舉個例子，比如之前軍人干政就有一個幹字。比如在文脈短的情況下，可能說幹這不像是人間的食物，可是同樣都有我們認為是髒話的這個詞，那這個情況不知道常不常見，或有沒有有趣的例子？

魏志平：

情感詞典是我請我的助理去看了幾百篇，幫我勾出來哪些是講情感。例子長這樣。有些字的意義在不同脈絡下都是有同樣的概念，比如說「滿意」，比如說「很好」。但確實有些字在不同脈絡下，可能意義上不太一樣。比如說「溫和」，在政治的場域上，溫和這個字搞不好不是太好的評論，說你是個很溫和的政治人物可能不是一個好的意思，可是如果你在家裡是個溫和的老公、溫和的太太，這就是一個大家稱羨的家庭。搞不好某些詞在某些情境底下確實是如此，那我回到情感字典，這件事情理論上都可以被取出來，我覺得機器有它的極限。人有時候都判斷不出來，更何況機器，再用更多文本去訓練它可能也沒辦法做到我們希望它做到的那麼準。

吳全峰：

請問比如以化粧品來講的話，溫和特定在清潔力下面的意涵嗎？還是溫和這個字比如說出現在質地或膚感的話，會另外給它一個詞性的分析嗎？還是它就是定型一個詞性分析？

魏志平：

做技術的人只能列一個正面負面的等第，我將之歸類，是因為我這樣比較能夠知道他在所指回想。不過很多字，即便在同個種類面對不同的面向，搞不好意思上不太一樣，如果要做到更細，應該要做到你剛講的這個面向。

吳全峰：

所以在分的時候並沒有清潔力、質地或膚感，這樣的分類。

魏志平：

沒有。

（四） 概念層級的情感分析技術

如果以技術層面相對是容易的，有些人會想辦法建立情緒詞典，如同上，譬如是分析旅館、分析醫療服務，利用該情緒詞典開始掃描每一個文本中出現的情緒的平率。正面情緒的字有多少，負面情緒的字有多少，當然也要考慮所謂的否定詞，如棒與很棒之間應要在分類，甚至可以利用其出現的段落給予不同的權重。利用一個非監督式的方法，想辦法計算情緒正負的分數，利用給予判定情緒整體的區分，這是以文件層級，相對而言技術層面較不複雜。如果要做到概念層級，邏輯上，要先找到產品特徵，不管是物質上或商業上的產品、被評論對象（如政治人物）、公共政策等，如附圖，先找出明黃色，在附近找可能針對該明黃色的產品特徵表達任何情緒。譬如娛樂選擇，在附近找到較少（於此為負面詞），那我們可以認定此為負面的。如食物旁有很棒，工作人員附近有親切，所以我們可得知為正面或負面。這個任務變成會有兩個子任務，一個是子任務萃取產品特徵，另外一個是透過人書寫習慣，在附近尋找有無含有情感態度的詞，我們就此可作標註。他可以是非監督式的方法，或監督式的方法。

稍微解釋這兩個詞，所為的非監督室是沒有任何訓練，透過衡量方法或是規則想辦法將我需要萃取出的東西決定出，故非監督式的方法以此情感分析的技術而言，針對人於寫作常用的習慣來訂定處理原則，利用這些自然語言的工具進行上述兩件子任務，萃取產品特徵、判斷文件中產品特徵的情感態度。第一個部分是，我們一般在描述產品特徵，決大部分是以名詞描述，第二個是，我們聚焦在較常出現名詞與複合名詞上，邏輯上是，你給予評論可能與產品、醫療院所、醫生服務品質有關，我們針對評論中的每句話作詞性標註，這是用程式去做的，留下名詞或複合名詞，譬如服務品質（service quality）就是複合名詞，成本（cost）就是名詞，我們就開始計算所謂常出現的指標，故我們希望留下明黃色這些字。接下來在判斷不同的產品特徵上，如果有一個以建立好的情感詞典，就在剛所找到可能為產品特徵的附近去找，有無列表中出現的情緒詞典，如果有，按照列表中詞典所標註的正、負，考量否定詞，決定該評論講到這件事到底是正面評價還是負面評價。這樣的做法當然會有很多的問題及極限，第一個是情感詞典不會是完整的，但我們可以訓練深度學習，讓其判斷該字為正面或是負面，既然字都可以投影在空間上，應該也可以訓練一個模型，利用空間上的位置關係，據此表明該字為正或是負，這是第一個問題，利用非監督式的方法，通常在實務、研究上不會有完整的情緒詞典，要想辦法利用其

他資源做語意分析，或是利用深度學習學習何字為正評或負評。第二個是，很多時候是使用隱喻的方式，譬如這支手機無法放進我的口袋，請問這是在形容何產品特徵？應該是尺寸，請問是正面還是負面？應是負面。產品評論非如同法律評論，會有較多的隱喻，這是較特殊狀況，惟資料量豐富時，如果剛剛提到得情況不是大宗的案例的話，在管理的研究上某部分是可以被忽略的。

（五） 以深度學習為架構之概念層級的情感分析技術

現在的做法是使用深度學習，下圖為類神經結構，我並沒有要介紹此，因為太複雜。此最終要標註的是，一個句子中哪個字是產品特徵，情感態度為何。最底層是 BERT，利用 google 讀了幾億篇上下文關係，將句子、字投影至空間上的模型。故簡單而言經過第一層 BERT model 後，每一個字、句子都已經在空間上的某個特定位置，接下來想辦法讓它標註中間哪個字、是否會產品特徵的一部份，到了 LSC，是在標註哪些字是正負詞，最終將二者合一，一起訓練，即得到標註結果。現在邏輯是用此複雜的結構，我是要告訴大家幾個技術上議題，這些結構非常複雜，很多的網絡、權重串連，以這張圖各位可以猜測有多少的權重需要訊量，可能是幾百萬個，現在以深度學習做應用分析的研究者都希望資料量越大，故標註資料就是現在做人工智慧中最辛苦的工作。

邱文聰：

所以利用 google 學習幾億篇下上文關係是人工標註？

魏志平：

Google 在學習並未做人工標註，他學上下文的任務是，給兩句話要判斷是否為同個文章中出現的，故他的標註是自動標註，以訓練文本模型而言。Google 有人工標註是利用大眾在做訓練，譬如我們使用 Google 有時會需要認證是否為機器人，會給予一些圖需要點擊，是否為人行道、紅路燈等，其實是在訓練 image。在做資訊的人會發現與很多研究領域不同，如果以大宗而言，學術包括美國的幾個知名學校，無法超越 google，因為資源太豐富。故訓練 BERT，全台灣沒有人辦法，將如此大量的數據，因為需要的 GPU 不夠，需要幾百個、幾千個 TPU。我要說的是 Google 沒有做人工標註，他的任務簡單，因為只想學句子意義，故他找出一個簡單的方式是，兩個句子如果在同一篇文章，要標註是否為上下文，類神經的訓練是要認出兩句話是否為上下文，認出上下文就有辦法學到是否為語言模型。一個簡單的例子是，他們最早是訓練字、克漏字，挖掉這些字，譬如我今天很__來到中研院，當你們文本是幾億篇或是幾十億的句子，前面會出現我今天、你明天，就會發現有些描述情緒的字會在一起，學出來的代表開心與不開心會在附近，因為他學校前後文的旁邊字可能差不多，

故將這些字堆在一起，這堆在一起是有意義的因為這些都在描述我。再譬如，台北前後出現的字與東京前後出現的字結構差不多，故城市（尤其是首都）都會聚集在一起。假設是有意義地抓到某些關聯，就可以計算、比對。

陳弘儒：

所以 BERT 不太透過字詞語意，而是算字詞位置？

魏志平：

他是用大量文本中，兩個字如果常在一起，而有關聯，他就會找出它應有的位置。就像 google translation 是用機率計算，有時候會發現翻譯結果很糟，可能是因為那部分讀得不多，機率較低，不像我們在做翻譯模型，需要很多配對樣本學習。完全就是當量大時，兩個字出現的頻率很高，前後文相像，這兩個字有關聯，這件事就是可靠的。

吳全峰：

所以 BERT 在這整個模型下，第一個所做的工作為何？

魏志平：

第一個 BERT 就是決定每個字空間上位置，同時參考上下文，這個 BERT 是 64 層或 32 層的網絡，非常深度的網絡，他的邏輯是經過幾億篇、幾十億或幾百億的訓練，他知道權重，所謂權重就是這個字與別的字的關係，故假設 AA 航空的服務品質沒有我想像中的高，我就將每個字輸入，每個字在原來空間就有其應該有的位置，接下來放進 BERT 中，他就會參考左右的字，在已經訓練好的模型，幫你把每一個字參考左右後，就會放進一個可能有上下文語意的位置，同時也產生這句話的代表的向量，這對我們而言所代表的向量就是這句話在空間上的位置。

黃詩淳：

如在傳統模型與 BERT 模型，如同使用同一批資料，最後成果差異為何。我想分享一個經驗，最近在參加一個日本比賽，邀請我們回答以前日本司法考題，訓練機器答題，視其正確率，各個不同研究者用了不同方法訓練模型，最後我發現 BERT 的準確度非常厲害，這讓我十分驚訝可以與傳統表現差這麼多。

魏志平：

如果問題本身不複雜，譬如只是分析這個文章是正評或負評。你的任務是答對是非題或是選擇題，這顯然較複雜，因為他不可能理解。如果是剛剛所說的只是要分正、負，較簡單的問題，我們做過的實驗用 BERT 有比較好一點，大概就是 1%-2%的差距。實務上，需要訓練這麼多資料、買這麼多 GPU，跟我直接使用關鍵字，如果只差 1%-2%，企業會認為不要如此複雜。有些問題確實是很難用傳統方式，譬如剛剛所提到語言翻譯，會有很多複雜關係，我覺得用深度學習應是一好方法。又如機器閱讀理解，輸入幾千篇文章、訓練樣本，用規則寫不容易，越複雜的是用深度學習較好，傳統分類有些時候會看到較大的差距。

這大概是深度學習架構的概念，很多人現在在用 BERT，這是一個非常複雜的語言模型。下圖的結果並非我們自己標註，是在做情感分析的公開資料集合，F1 大家可以想成是準確度與召回率的平均值，我們抓出這個字為情感的字，態度為正確或負面，完全正確我們才算正確，F1 可以到 67-72%，我認為在這個資料量下還可以接受，所以這是利用複雜模型訓練出的結果。

五、使用者輪廓分析(user profiling)

（一） 使用者輪廓分析重要應用

在使用者輪廓分析中，想辦法從一個人留下的文本輪廓出這個人的個性，譬如可以推估使用者的人口統計變量，是男或女寫的、年齡層為何、教育程度為何、人格特質、政治傾向在台灣可能可以很清楚的辨識，在美國也是。如果剛剛所說的性別從文本推估，通常被視為文件分類問題，過去大概是用文字作分析，去年有一組學生是用 Instagram 的圖片做分析人格特質，準確度還可以。User profile 最傳統是用文字分析，我認為圖片應該是下一波，這件事是否有重要運用，如果可以知道男生、女生對於產品或是政治人物的評價是否有所不同，年齡層對於某些產品不滿意的原因可能會不同，可以有更細緻的分析結果。也可以協助做產品推薦，現在的產品推薦很多是以過往點擊的產品在做推薦，但很多傳統推薦其實是在人口統計變量上，很多的推薦可能因為對顧客不了解，而是用概念做推薦機制，如果能夠針對使用者的 profile 分析，對於電商網絡平台做產品推薦可以增加額外推薦方式多元性。最後還是講到機器人，聊天機器人如果知道聊天對象可能是男生或女生，可能的年齡層，使用的文字上、話語的標註會有更多元的選擇，做出適性化、個人化的回應。

（二） 範例：更精細的網路輿情分析

我針對某航空公司大概兩至三百篇評論的統計，如果給這麼大約的分數，對航空公司而言只會知道可能還不夠好，如果可以給更細緻的分析，較能更符合顧客要求。有這些細緻分析後，網民意見的看法能夠有更精細的討論。

（三） 使用者特徵影響寫作風格與用詞？

這如何做 profile，就社會學或語言學的角度觀之，使用者特徵會影響到寫字的風格，如 Schler 等學者 2006 年分析 blogger.com 超過 71,000 部落格內容發現，不同性別與年紀層的寫作風格有明顯差異。男生年紀較大的部落客使用較多的介系詞，女性、較年輕的部落客較常使用代名詞及贊成或否定的字詞較高。而 Otterbacker（2010）學者提到女生在寫評論時傾向用第一人稱，男生評論較常使用第三人稱。Rao 等學者（2010）的研究發現年輕的推特使用者較常使用重複的字詞來強調他們的情緒。Schwartz 等學者（2013）分析 75,000 臉書使用者的 15,400,000 的臉書貼文發現，女性使用者的臉書貼文較常使用情緒字詞(如“excited”) 以及第一人稱單數，而男性使用者的臉書貼文較常出現髒話(swear words) 。外向的臉書使用者較常使用一些社交相關的字詞(如“party”, “love you”)，而內向的使用者較常使用與個 人活動有關的字詞(如“computer”, “reading”) 。比較神經質(neuroticism)的臉書使用者較常使用“sick of”來表達他們的負面情緒，而高情緒穩定的使用者比 較常描述令人愉快的社交活動之字詞(如“sports”, “vacation”, “beach”) 。

（四） 可以用以推測使用者輪廓的變數

綜合剛剛所說的文獻，我們建立一個推估使用者輪廓的變數，依是文章中使用的字詞，有些字詞較容易出現在男生或是某些特定的人，挑出具有區隔力的 300 個字。寫作風格考量，譬如詞性，男生一般而言比較不會用太多的副詞，女生較會使用，在詞性的分佈比率不一樣，在功能詞性，包括介系詞、代名詞、限定詞、連接詞、助動詞及介副詞的使用可能不一樣，因此我們採用 6 個變數，停用詞的比率、詞彙的豐富度，即同文章中不重複的字詞佔所有字詞的比率，文章中句子的長度，男生的句子通常較短，女生的句子的修飾程序較大，字數的統計上會較多。

（五） 使用者輪廓分析實驗

用這麼多變數建立模型去做實驗，我們使用較有強烈性別傾向的產品，例如刮鬍刀應該是男生會購買，女裝、女鞋、瑜珈褲應該是女性會購買，故我們至 Amazon 的網站上搜尋這些產品的所有評論，暫時標註評論這些產品的是女生或男生，當然也無法排除幫老公、老婆、兒子、女兒購買等，這些就無法得知，故在評論中出現有 boyfriend、girlfriend、husband、son、wife 等字詞的評論刪掉。至於 Airbnb 並非用代號，他的名稱通常是純英文字，故我們根據評論者的人名，將放置到 Genderize 的網站，Genderize 提供的 API 來標註評論者的性別，這個 API 根據提供的名字回傳最可能的性別以及機率，最後我們只保留那些性別機率在 0.9 以上的評論。

另外，我們找了兩個評論網站 Yelp、Tripadvisor 做為資料來源，我們想做的是國家比，因為這些網站都有提供評論者的國籍資料，假設國籍資料是正確的。Amazon 是 62,000 筆的評論，Airbnb 是 54,000 筆評論，國家別比較少分別是 6,400 與 7,800 筆評論，這是使用我們的模型跑出的資料，平均而言準確度可以到八成。性別 Amazon 的數據較準確，Airbmb 大概 70%準確。簡而言之，我相信當我們用更複雜的深度學習的模型，此模型僅有用 vector space model，如果資料量更大用更複雜的模型，或許準確度可以更高。這是 user of profile，這個議題會與個人隱私或是影響更大，當一個網站握有你許多資料，影響層級可能更多，不單純只是商品推薦的準確度，可能影響到政治傾向與性別傾向等。

六、品牌個性偵測(brand personality detection)

（一） 品牌個性

所謂品牌個性，也就是以人類個性特徵來個性化品牌，如「哈雷」給人的感覺就是粗獷，「維多利亞秘密」給人的感覺就是性感、漂亮的等等。在管理上，企業的品牌經理人基本上會透過分析可能的市場或定位，形塑其品牌個性。如麥當勞用廣告創造歡樂形象，例如世足賽、奧運等行銷。很多研究認為品牌個性與人的真實個性是吻合的，他的品牌態度、品牌偏好、品牌忠誠度和信任度、購買意向與正面口碑會較高。故形塑品牌個性是企業中在建立品牌很重要的一環，重點是企業想要的品牌個性與消費者真實感受到的品牌個性可能有所不同，所以品牌經理人需要不斷確認企業想要的品牌個性與顧客想要的是否有落差，因此需要定期發問卷詢問，確認企業目標的評排個性與顧客感知的是否一致，甚至是不同年齡層、文化對於同一品牌的感受。

（二） 品牌個性管理的重要工作

這些都透過問卷詢問將會曠日費時、成本高，尤其是一個事件發生後，企業經理人還要評估該事件發生後對於品牌個性的影響，以及危機復原的程度到何種程度，如果都以問卷，事後可使用，事前卻沒辦法，下圖的資料是 BrandZ consulting firm 在全球做的品牌 Volkswagan，兩個品牌面向值得信賴度拿到 25 分，不誠實程度不高，大概佔了 8%。過去 2010~2015 年的全球調查都是差不多分數，到了 2015 年、2016 年，發上了排放的醜聞，2016 年的調查兩個品牌面向都各自下降，經過一年調查才逆轉。故企業在看品牌個性，遇到危機或是損害事件會採取災害復原，要先知道受損有多大，經過半年的災害復原，譬如推出免費保證等，能不能在不同的品牌回到原有水準之上，這些都是重要的。在企業上一直用問卷詢問的成本太高，有沒有替代方法，我們的研究就是做這件事。

（三） 以社群媒體資料為基礎的品牌個性偵測技術

我們視為分類問題或是回歸問題，如果品牌個性按中間值分成高低就是分類問題，如果以 0~100 分就是一個回歸問題等。分類以傳統方式而言，可以

將重要字詞挑出，簡單而言該品牌可能有 1000 篇評論，我們就將此 1000 篇評論當作是一整篇的文章，我們就將文章中最重要的關鍵字詞，就權重找出當作舉證，據此建模型。也可以使用深度學習即 BERT，找到每個句子或文章空間上的位置，再把所有與該品牌有關的空間上的向量取出平均值，即代表該品牌所有評論對他的觀點，據此建模型。簡單而言，我們認為品牌的值得信賴度高者會出現某些特定的字詞，值得信賴度低者會出現某些字詞或用語，技術上大概是這樣。

（四） 品牌個性偵測的評估實驗

我與在數據庫的學生合作，歸納 20 個正面的不同品牌個性，樣本大概 819 個，涵蓋了航空公司、汽車、咖啡、尿布等 196 個品牌，我們擷取 twitter 的數據，大概兩百萬個 twitter，平均而言 2,400 篇評論，這是做出的結果。回到剛剛詩淳老師所提問的，如果使用傳統的 Vetor space model，我只考慮關鍵字，不考慮關鍵字中間的意義、關聯，使用兩個顏色區別，超過八成使用藍色，低於七成使用紅色，在最中間那一欄大部分是低過七成。如使用 BERT 為基準，大概是多 5~6%，在不同面向都贏過傳統技術，平均而言我們將 20 個品牌個性準確度提出，傳統只考慮字詞，可以到將近七成的準確度，如果使用 BERT 可以將近 76.28%，大概多了 7%的準確度，這個模型還是有她中意的價值存在。

七、未來展望

最後，仍需要克服許多技術、管理的挑戰。從技術層面而言，第一個我們一直在想辦法提升準確度。第二個是，如果以產品分析、情感特徵大家使用不同字詞描述相同或類似的東西，譬如價格、房價、CP 值，我們要想辦法類似的歸類在一起，使能夠彙整不同人寫作的內容得到一個合計後的結果，這件事必須要做到。再者是話語，如果是單一國家的品牌，可以以單一國家製作。惟像是航空公司，評論者來自不同國家、不同語言，如何合計、分析不同語言。此外開發與改善偵測假評論與假評論作者技術，譬如我們做過中國電影評論網的分析，在中國電影評論的寫手很多，因為電影可能成本小，需要仰賴寫手幫助，不只有中國，可能很多地方也是。假評論本身應有些線索可得而知，如細節部分無法完善，字詞通常會較模糊。假評論是有時間限制，不能等到產品出來再寫，通常於適應期就大量寫評論，這些人的評論通常會是極端正評或是極端負評，不會給予 3、4 分，對於影響評分沒有幫助。故會有內容上的特徵，可以辨識出，評論的行為本身是有線索的。最後是利用大數據的實證研究與實務意涵的討論，應審慎地檢視在社群媒體發言的樣本與所有顧客的樣本是否存有很大的差異，社群媒體上的樣本如同我們認為鄉民有辦法代替全部人民的想法嗎，可能無法。社群網站發言者不管評論產品或其他，他的樣本與所有顧客是否有差距，這是利用大數據做相關實證研究在探討實務意涵時可能要特別關注的問題，惟對做技術的人而言可能會是下個運用端的人要去思考的問題。

問題討論（敬稱省略）

蔡政宏：

前面有關股票漲跌與政治輿情時，事實上與假評論類似，但可能不是假的，而是需帶領某個風向，類此情況。我認為自然科學相對是較客觀的，較極端如同假評論，在輿論或股票操作，某些人想要達到目的可以透過此達成。第二個，情感詞典有提到 Metaphor 或是我們會說言外之音是較偵測不出的。我好奇的是像在語言哲學或是語言學中提到人在溝通時，我們會遵守幾個溝通原則，例如今天晚上看電影，另一人回答功課未完成。事實上兩者對話無關，故打破溝通相關連的原則。有無可能透過語言學或是語言哲學，後面有提到聊天機器人要找 profile，最後還是需要知道主人的情緒為何，剛剛的情境中一方是不想去，所以只是看對話內容而不了解情感，這部分有無可能解決。

魏志平：

我先回答最後一個問題，我於 1988 年碩士二年結束就開始研究人工智慧，在處理文字時，會一直回想字詞上的意涵。人工智慧在處理文字這件事，過度受到複雜東西的影響，到了深度學習的時代，沒有任何語言學的基礎，我們做自然語言處理會認為詞是一個很重要的概念，譬如 computer center，單說 computer 或 center 可能不知所指為何，合在一起就可知道所指為資訊中心，詞是有意義的，將多個詞按照某次序組合，就有語言上的意義。他們在輸入 BERT 時是使用單字輸入的，以解決程序上的複雜度，這會有一個問題是 out of vocabulary，因為總有可能有沒看過的字。最後就拆字，拆成更小的單元，以解決 out of vocabulary 的問題。這可能會變成語言好像比字還小的 component，我認為 BERT 的邏輯不遵循語言學的概念，沒有所謂語言分層好幾個層次的概念，他是用大量數據訓練，他認為要能夠運算一定要放在空間上，這個字的前後常出現某些字，那這個字放在此應是正確的位置，經過前後上下文的判斷就可以將句子放進適當的位置，他是使用極大量的數據想辦法 imbedding，在空間上找位置代表他。

吳舜文：

如果其他領域研究者想要利用 social media 分析，例如今天我想做一個政策的分析，有沒有軟體可以使用，還是一定會需要寫程式？

魏志平：

我們程式都是自己寫的。

邱文聰：

回到蔡老師的問題，現在這種應該不會稱為自然語言，已經拆成看不出是語言，只是透過資料算出位置。也因此很難期待真的讀得懂玄外之音，譬如棒、好棒、好棒棒，可能會有不一樣甚至相反意思，很難用拆解字位置就真的有辦法抓到微小的差異，他有很厲害的地方，但也恐怕有它的極限。

魏志平：

我覺得任何技術都有侷限，有些單位特別是企業界都可能會高估其準確度。管理學院都會考量成本、效率概念，機器能做的是大量這件事，不見得很準確，他能夠讀 2,500 萬份文獻，人終其一生可能也無法達到，機器可以在幾天內分析完，或許有一定比例的錯誤，但當這麼大量時一定要靠機器。在做技術的人常常會有一個 Ground Truth，經過較嚴謹的訓練訓練機器做準確度到八成，同樣的東西找人做，做出來的答案相同，就認為超越人類、人的極限。為何人類會輸給機器，應是人的會疲倦、前後標準不一致，如同改考卷，任何兩份我的評分標準，會經過再檢視，會希望把差異降低。人會輸給機器我會認為可能是因為認知負擔或疲憊，不是因為人真的做不了這些東西。期望或許再過幾年玄外之音有辦法被突破，當看過更多文本後。

陳弘儒：

做此分析時是用很大量的數據進行，在管理學中，是否有朝著不用太多數據可以做出相同的分析的方向前進。且某個程度上文本分析是可以無限的。

魏志平：

在技術上有所謂 Transfer-limited，很多不同層次的移轉，簡單而言，如同剛剛的例子，想要分男性、女性的評論，可否拿 Amazon 的數據當作訓練樣本，真實的目標可能是分析 Airbnb，其中的寫作內容可能不一樣，因為用戶的特性不同，寫作的內容、風格不一樣，這時我拿這個東西學習，某種程度上利用此更新，其實人有這個能力。人在看某些東西，再以此為基礎看類似的東西會有基本的理解，這叫做 Transfer-limited。我們今年剛作完閱讀理解，才標註 3 千多份數據，是由台達電放出一批數據，是在做一般性的閱讀理解，我的目標就是標這麼多，我想做醫學的對話，我們發現確實可以轉換，在抓題目跟答案中間可能還是有基本的邏輯變化，這叫做閱讀理解。技術上很多人想辦法再突破，即能不能將無關的是納入。

莊芸芸：

老師請問一下，我們都是到機器學習耗電要求機器設備，如果需要使用 BERT 模型需要有什麼設備才可以運作此模型？

魏志平：

大概要幾個 GPU，一個大概四萬到五萬間，因為 BERT 十分耗運算資源。

莊芸芸：

假設 BERT 模型退下，老師有推薦什麼樣的模型替代？

魏志平：

BERT light 是更小的 BERT，我們其實都用 BERT light。我認為跨領域合作才有辦法，讓我們去學法律很困難，因為邏輯思維完全不同。

李建良：

現在在檢索系統輸入字詞進去，以前會是與檢索系統一樣的才會跳出，但是現在會將字詞拆開，顯示一樣與相關的。

魏志平：

早期是建立字詞關係，人工智慧算其他關係性給予權重，現在我相信也適用字詞的空間位置運算。

李建良：

這種發展是我認為進步，在檢索時會檢索到許多不一定需要字詞一樣的資料。再者是，剛剛提到假評論的問題，其實會有三種情況，譬如影評，可以給予寫手看劇情內容，寫假的評論。第二種是他是盲目，沒看就寫評論。這兩種情況會不同，如果是希望能帶風向可以用第一種情況。而第三種情況可能是用人工智慧寫評論，將來可能寫手根本不是人，我們從閱聽者角度會有這個問題，另外是對做網絡媒體分析的人，也就是要分辨這三種情況。

魏志平：

現在一些網站是用管理機制限制，完全排除很難，譬如 Amazon 要購買才能夠評論，一部份的盲目的人會被排除。只是後來發現，如果真的是公司可能也不在意購買，只是讓中介商得利，正評還是會增加。故某幾個網站進一步規範，只能寫買過，一個月最多只能寫兩個評論，再削弱寫手被操作的可能性，技術一定要偵測到哪些是假的，讓網站的人有辦法找到有嫌疑的人進行調查，也是在解決控管上的問題。

李建良：

在網路世界會變成資料量是十分大，問題才開始，這是一個我關心的問題。剛提到跨領域研究，這我非常同意，也是請魏老師演講的主要目的之一，希望未來程式

第三次會議紀錄	
時間	109 年 9 月 29 日（星期二）
主題	人工道德與人類直覺
講者	蔡政宏（中央研究院歐美研究所研究員）

內容摘要

壹、哲學與 AI

人們對於哲學的理解大多停在舊時代，如亞里斯多德，較近的會提到的也是 Decarters，但也是接近四百年前的學者，有時討論 AI 時會認為 AI 與哲學並未有太大的關係。我們翻閱 AI 的教科書會發現，清大，目錄的最後部分有討論的哲學問題，特別是強 AI 與弱 AI、倫理議題部分都是他們特別關心的。哲學家長期以來都很關心心智是如何運作的，機器如果可以跟人一樣行為上有智能展現，可以透過機器如何運作回過頭理解人類心靈的運作，或是透過心靈如何運作建構 AI 是如何運作。這部分是哲學家所關心的，非單純是技術問題，關心的是對人自己的關懷。在跨領域與不同學界交流時，發現大家使用弱 AI 與強 AI 字詞的意思都不一樣。在哲學領域中，弱 AI 是指，他的行為是如同人類一樣有智能，該機器並非有心靈。強 AI 是，該機器與人一樣有心靈，我們一般聽到 AGI、ANI 在哲學中的劃分都是屬於弱 AI，機器執行單一任務或多項任務，在我們看來都是機器仿造人類做法的智能行為，但是他本身並無智能。

John R. Searle，提出中文房論證(Chinese Room Argument)，是為了挑戰 Turing machine 的架構。Turing machine 的運作是，透過一個房間中有個機器和一個人，使用者透過對話並且不知其中為電腦或是人，如果該使用者

無法區辨，該機器就可宣稱有人的智能。John R. Searle 就是要挑戰這樣的假設，中文防盾正就是，即使通過了這樣的測試，裡面的人也是不具有智能的。但仍獲得許多回應，其中一個是並非其中不懂中文的人不懂中文，而是以整個中文坊看其是否懂中文，不能僅看其中的人。而後來有許多種回應，不在這裡討論。哲學家所關心的是智能運作為何，ANI、AGI 在哲學領域看來較屬於弱 AI，有時在與他人溝通會特別強調對方所指的為強 AI 抑或是弱 AI。

另外是 Dreyfus，也是反對強 AI，他認為電腦或是機器無法擁有人類的心智，他們論證方法與 John R. Searle 不同。透過研究人類展現技能(skill)、專業(expertise)、知識(knowhow)等，這幾個字對他而言是同義詞，各位可以設想當真的掌握某種技巧，且技巧的記憶是很高超時，如同庖丁解牛一般，不用思考內部的規則，如開車完全與車子融為一體，在這個情況下沒有規則。Dreyfus 透過技能等考量看出人類的心智，再對比人工智能，至少在他的時代，rule-based、強調專家系統的時代，他認為有根本差異。他有幾本書，強調人類的直覺。這部分他將人類 skill 分為五個層級，越低的層級越會用到規則，如設想在開車時，新手開車時教練會提醒很多規則、技巧，當技巧從新手到越高等級時，越高等級的技巧擁有者所用的就不是規則，而是直覺。這部分也有許多討論，如做決定時，Dreyfus 有時分成五個階段、有時六個，最高等級為 wisdom，無論如何初步來看大步都是有五個階段，即技巧學習有五個階段，最高階段都是用直覺做判斷，直覺的判斷是不用將要處理的情境劃分成不同的單元逐步的分析，而是直覺不用思索就知道如何行動。Dreyfus 認為專家遇到情況可以不用思索，直覺就可以看出。如果一個專家在做一個行為表示，並不會有下列心靈狀態的運作，有些用紅色標註。早期在談人類心智運作，大概從黑格爾以前會較強調理性的重要性，在二十世紀有些關於理性的爭執，開始質疑人類的理性，理性的重要性在哲學界開始產生爭議。這裡在談的是，在直覺狀態下根本不必用到思慮，不用意識控管，甚至不用費力、沒有自我，如同無我或是全神貫注的狀態，心理學所說的「flow」的狀態。甚至在展演這些表現時，如果還用到心智會干擾到表現的好壞，這是對於直覺較多的描述。

在哲學以外心理學有很多研究，經濟學家也是，這部分的直覺一般而言他們稱作：Expert intuition，有許多的研究。有些是在實驗室中，有些透過實際情況，例如考察消防隊員，消防隊員在做抉擇時常常沒有時間做判斷，或是飛機駕駛員等是如何做判斷，這些他們稱為自然主義式的決策，不在實驗室中觀察，而是在實際情況觀察。哲學家對 AI 關注的是，人類與 AI 對比時，從人類智能、道德、創意、意識或智慧，討戰 AI 是否有這樣的智能、道德等嗎，在挑戰時會回過頭來看，原本我們認為是所謂的智能或道德、創意，也許與我們的想像不同，如果 AI 也可以展現的話。如果用人類的看法挑戰 AI，有時回過頭來看想法可能會鬆動對於人類原先自我認識的看法。較大的圖像可能是，在做此研究時，可能是兩邊相互的衝擊。我較關心的是人工智慧如何可能，為何

人工智能要到人工智慧的階段，到人工智慧的階段後會遇到理論上的挑戰，很可能是原則上的不可能，因為智慧者與 intelligence agent 的思維模式不同，智能者無法做出人類智慧的思維模式，如原則上可能必須滿足何種條件。假定原則上可能，實際上真的可能嗎？這是我想要了解的部分。後面的部分會談及從人工道德在實踐上如何可能，哲學家問如何可能時，並非認為完全不可能，而是大部分都會遇到理論上的阻礙，於此理論上的阻礙即仿人類道德的設計道德，可能有兩個方法。一個是由上而下的方法（top-down approach），指定或貢獻道德規則給 AI。另外一種是由下而上（bottom-up approach），讓 AI 自己學習。但這兩種方法可能會遇到某種挑戰。從此處我會在談及人工智慧是否比較不會遇到此問題，其實也是會，故我會談到有關 well-being 某些資訊會衝突。

問題與討論（敬稱省略）

邱文聰：

人工「智慧」的智慧概念的內涵似乎與福祉有密切的關係，可否說明智慧為何以福祉作為概念核心？

蔡政宏：

這是一個有點複雜的問題，我訴諸哲學家較可能接受的定義關於智慧，其實這個問題十分麻煩，哲學家自己也無法通透，有很多種講法，我所採用的是 Green 2015 年現在聚焦、比較能接受的看法，這只是一個必要條件。一個主體是實踐智慧，智慧有時會說理論智慧，我們說的大部分是實踐智慧，可以說是個定義、必要條件之一。如該主體知道何 well-being 對他而言是重要的，可以當作規約定義，如果不將此當作規約定義，換個方式，也可以不用智慧這個詞。人類如果想要獲得 well-being，獲得 well-being 有兩種可能性，一個是碰巧得到，另一個是可以培養某種傾向，讓他穩固、穩定的獲得 well-being。而所謂穩固、穩定的傾象，有人會習慣用 virtue，即德性，如果擁有某種德性，即可以獲得某種幸福。

這裡用的是智慧，如果不使用該詞，可以換一個詞。這部分，一部份會訴諸規約定義，一部份是可以不要管智慧這個詞，我們直接問人類有無一個穩定傾向，這傾向可以使你獲得 well-being。well-being 可能是人類活在世上最主要追求的，論證可能是，如為何要考大學，是為了一個好的職業，為何要好的職業，是為了要有好的生活，會無限後退，只要給一個回答就可繼續問。回答到最後，為何要追求詢問，所以有時 well-being 問到不能再問人類在追求最終的東西。關於何謂 well-being 還有很多討論空間，這裡有好幾套理論會稍微帶過。有人認為 well-being 是種愉快，這裡我將智慧與福祉，是透過剛剛的條件連結，

也就是一個 agent 如果是有智慧的，他知道什麼對 well-being 最重要。

【續上報告】

AI 的教科書中常常會提到，大部分的 AI 研究者 program 能否運作，事實上不關心 AI 是否真有智能、道德，我認為這部分需要很小心。機器做出，例如可否跟機器人結婚，此問題可能已經超過機器構造層面的問題，或是對於自主機器人可否咎責，如醫療機器人，或是機器人是否可具有公民身份。這裏特別關注的是，如醫療機器人會分等級 0～5 等級，關於機器人最後可能會有有高度自主性、完全自動化，在人類世界活動時，與人類互動與人一樣，也可以單獨決策，在此情況下，就要明確思考關於責任問題，如何定義該機器，這些哲學問題並非不用思考。

貳、AI 機器人

AI 機器人能做什麼、應做什麼、與規範性的關係為何。有許多問題值得我們關心，如機器人可以自由進出主人家中的所有房間嗎？這可能已經超越機器有無完成任務的問題。機器人可以聽從主人命令去鄰居家拿他人的手機嗎？或是機器人面對男主人命令他去泡茶，女主人命令他去洗碗的指令應如何回應？或是機器人將晚餐倒在地板上，我們可以對他大吼大叫嗎？這些問題看起來一般，我們可以將之延伸。例如為何可以進出所有房間，特別是在該機器人有連線可以截取所有影像上傳雲端，這時就要考慮到此類簡單問題外的議題。至於聽從主人命令是要考慮的是，是否可辨識出主人要求其執行不道德行為，不管嚴重與否，機器人可否辨識出。所謂泡茶、洗碗看起來很簡單的事，但我們疑問的是機器人可否判斷哪個人的命令具有特別凌駕性。最後是要考慮機器人採取某種態度，是視為人還是視為物，這些可能已經超越到設計層面問題。

與 AI 相關的規範，有時談及規範關心的層面可能不一樣，可以分成三大部分。一個是製造者相關的規範，有的是關於使用者相關的規範，有個是機器人本身需要遵循的規範，某些書籍中是做此分類。我有些不理解，前面並非針對機器人本身，而是針對機器人製造者、使用者都放進 Robot Ethics 研究，關於 AI 機器人本身需要內建何種規則或道德規範，稱作 Machine Ethics 放在該領域中研究，我們在討論到機器人時特別都會討論到所謂三法則。前面部分是關於製造者的規範，國外也有很多討論，主要是針對大公司在製造機器人時的相關規範，可能都是在做倫理漂白的工作，因為沒有約束力。故關於機器人的製作者，到底需要遵守何規範？法律效力為何？這是目前大家所關心的。我叫關心的是 AI 機器人本身需要遵守何種法則，應如何內建給機器人，機器人的如何思考的。大家可能常聽到會給機器人三個法則：1.機器人不得傷害人類 2.袖手旁觀使人類傷害 3.除非違反法則一，否則應遵從人類命令。

問題與討論（敬稱省略）

邱文聰：

這不算是對製造者的規範？

蔡政宏：

應該是說該法則適用的對象，或是遵守的對象、主體是 AI 機器人。

邱文聰：

看起來像是機器人，但最後是在規範製作者，必須製造出這種機器人，必須遵循這三個法則。

蔡政宏：

我們可以有兩個規範，機器製造者應將好的道德規範歸屬給機器人，這是一個規範。另一個規範是，機器人不得傷害人類。主體不一樣，法則或是規範可適用的對象不同。我了解文聰的意思，若此我會分成兩個層次，製造商遵守的層次是，必須將 X 法則給機器人。

參、AI 機器人「道德推理」：人工道德

一、理解問題

目前哲學界畫分幾種可能的做法，我們將機器訓練有道德，一種是由上而下的方式，人類的道德原則或將道德原則的理解寫成程式再給機器，這部分還是需要專業工程師。哲學家所談論的是相對抽象的，特別是道德原則，與道德原則相關的語句，具有應然與句的特徵，在這部分，透過應然邏輯，寫成程式，但這部分還是站在較抽象、原則性的方式撰寫。另外較不抽象的是，2018 年的，也是試圖使用邏輯方式將道德原則形式化，形式化後，後面工作仍須由工程師寫成程式。

我認為這還是非常抽象，因還 codes 要如何理解。接下來要討論什麼是傷害，對我們而言傷害此十分簡單的字眼，例如機器人不得傷害人類或是不得不作為使人類受到傷害，傷害要如何理解，假設 AI 機器人看到人類甲在追打人類乙，這時是否要判斷甲在傷害乙。如果從人的角度可以很快判斷，惟如情況特別時，一個警察在追打強匪，人類甲並不只是人類甲而具有某種身份，人類很容易辨識出，但是 AI 機器人有辦法很快辨識出嗎？在這可能要將法則 1 改寫成法則 1.1，警察對人類的傷害不算真的傷害，但又會卻步，如在香港的情況。故在此可能要一直修改，AI 機器人無法像人類一樣具有很大的基礎。還可以提出很多例子，例如人類法警在執行死刑時會有心理壓力，可否改由 AI 機

器人執行槍決，這時這個機器人是可以殺人的，無論理由是否接受，故可以延伸討論，可否殺人、槍決，甚至是執行安樂死的機器人。

可能的解決方式就如同上述，必須一直修改 AI 機器人內建法則，如 AI 機器人在特殊場域刑場、戰場、醫院，對人的傷害不算傷害，這是特例，惟那樣的傷害應如何理解、如何形構出。又例如性愛機器人、性虐待機器人也是在傷害人，可能必須不斷將傷害這個字眼的內涵修改。較擔心的是我們無法將所有情況窮盡，但是對人類而言可以馬上判斷是否為傷害，所以可能會反問人知道是否為傷害，對人而言，因為人有身體，像在設計機器人時，人有前後之分，有些行為會往前做不會往後做。但 AI 機器人非內建或是生來就有這樣的身體，身體既有某些型態，人類生來身體的型態會使人類往某個方向做，但機器人並不見得是這樣。故說到傷害時，為何人類可以很快理解，這部分很複雜，一般可能會認為是因為我們有同情、同理，甚至是因為我們可以透過同理知道這件事我們不喜歡。但 AI 機器人不同，他沒有受傷的感覺，我們可以很快推理，這個情況下我會有什麼感受，而這個感受是我不願的，所以可以知道這件事該做不該做。但 AI 沒有這些材料、沒有這樣的身體，幫助其做資訊處理。另外是訴諸對於文化世界的一般理解，如牙醫對病患造成的也只一種痛，但我們不會認為那是傷害。故我們在教導 AI 傷害與痛是不同的，外科醫生也可以對人體進行手術，但人類可以很快區別出那並非傷害。故必須不斷地將傷害的內涵增加，惟應如何增加是我們必須面臨的問題，特別是 AI 可以完全自主化時。故這是由上而下的方式可能會遇到的問題，到底應選擇何種規則就有很多爭議。即使選定，其中很多基本的字詞，如何將內涵寫出，這也是問題。在此情況下，有人建議不要給規則，由下而上的方式學習道德。

二、習得問題&道德直覺

AI 機器人要學的道德推理如何運作，我們讓機器人自己學習，與人類小孩相同，並非一出生就告訴他應遵循的道德法則，而是透過情境學習，逐漸獲得道德法則。這些部分心理學家進行很多的工作，他們在想是否 AI 可以做這樣的練習，給予大量的案例，問題是不論是否為監控式的學習，回到人類本身而言，既使蒐集到人類本身大量的案例，代表的意涵為何。目前麻省理工就在搜集人類對於道德兩難時會做出的道德抉擇，最後做出統計，到底什麼是比較重要的，大家的選擇會較注重生命的數量，還是年輕人或老年人。在台灣也有版本正在進行，如清華大學人工智慧研發中心的丁川康老師有做過，目的創建倫理案例庫，他們考察的案例比麻省理工更多，除了自駕車外，如工程師為增加產品利潤，將產品壽命減少，工程師自己是否應舉報。他也將主角改成 AI 機器人，AI 機器人本身是否要舉報。又如 AI 護士照顧成人，病人不服藥會死亡，服藥會感到痛苦，機器人可否強迫病人服藥。

我的問題是，蒐集到的數據，應如何運用，蒐集到的是螢幕後的人的道德直覺判斷，還是深思熟慮下的理性判斷。另外是這些案例並未搜集到人類下判斷時的理由，故 AI 學習時也不會學到理由。也許在電腦螢幕後下判斷的這些

人，下判斷時，我們所要詢問的是，人類的道德判斷是如何形成的。我們參考Haidt，他是一位心理學家，研究人的道德判斷如何形成，他最近將道德心理學上的發現運用在美國政治圈，民主黨、共和黨的這些支持者，道德判斷形成判斷時是如何形成的。他的想法可以使用很簡單的標語看待，即形成道德判斷時，通常是直覺先行，事後講理由。也就是先有很強的直覺認為正確與否，至於理由非一開始就有，這與「人是理性的想法不同」。即我先有某些理由，再從理由推至這件事在道德上的對錯。他提出了幾個例子，我們在做直覺判斷時並未有理由，但在詢問他為何做這樣的判斷時，他會自動將理由補上。如：下面有四張卡片，每一張卡片有一面是英文字，一面是數字，規則是如果卡片的一面是母音的話，另一面就會是偶數。問題是要判斷該規則是真或假，必須翻開哪幾張卡片檢查？他是一個條件句，A 是母音，我們可以看另一面是否為偶數判斷。如果卡片一面不是偶數，則另一面一定不會是母音字母。這是心理學家在測試人是否有理性時，最喜歡測試人對條件句的適用，透過條件句判斷人是否有理性，這是他們常測試的其中一個版本。大量測試後發現，大部分會選 A、4，因為看到題目時會專注在母音及偶數上，事實上從邏輯角度要挑 A、7。直覺上會有某些判斷，Haidt 紀錄無論受試者聽到正確答案或是錯誤答案，這些受試者都可提出說明為何做這樣的選擇，且推理還顯得十分有自信。

另外，他做兩個道德上的測試。測試一：茱莉和馬克為姐弟，一起在法國旅行，單獨在海邊的小木屋用餐，如果這時做愛會更有趣，茱莉長期服用避孕藥，為保險起見馬克也使用了保險套，雙方都十分享受，但也認為不必要再嘗試，他們將當晚發生的事當作彼此間的秘密，這樣一來兩人的關係也更加親密，請問你對這件事有何看法？他們做愛是不道德的嗎？是否為亂倫？測試許多大學生？他發現很多人直覺認為是不道德的，因為會生出畸型兒，但故事中已經提出女方有服用避孕藥、男方有配戴保險套，不會有這個問題，受試者又會提出其他理由，實驗者一一駁斥，受試者啞口無言，但仍堅持這樣做是不道德的。

測試二：故事人物珍妮佛在醫院的病理實驗室工作，他認為殺動物是不對的，基於道德的考量選擇吃素。不過，某天晚上他必須把剛死的人類屍體火化，他認為將屍體丟棄很可惜，變割下一塊肉帶回家烹煮後食用，請問珍妮佛的做法是否不道德？或是可否這樣做？大家直覺為不道德的，當初理由是基於殺動物是不對，事實上動物在自然死亡，為何還是認為不能食用。且動物有生命，植物也有生命。於此要問的是我們下道德判斷時，需問反對與支持者的理由為何。H在此的結論，於測試一只有 20%認為可以，測試二只有 13%的受試者認為可以。實驗者有記載，他會不斷詢問反對的理由為何，他的結果是，在這些無害的禁忌情境中，受試者會提出很多反對理由，反對理由數量不斷被拋棄，遠多於其他的情境，有些情境較單純，如在這為何不能，會不斷提出理由，直覺先行、理由後援，因為理由一直不斷變化。

這是 Haidt 一直在討論的議題，最後提出一個社會直覺模型。一個觸發事

件開始時我們會形成某些直覺，好或不好，在直覺情況下會形成某些判斷，該判斷需形成理由，為何下此判斷，此理由根據作者想法，理由的目的是說服、影響別人，故該理由行程時，是為影響另一人 B 的直覺。但通常自己形成的理由不太會去糾正、挑戰自己，故理由不會回到自己身上，而是以該理由更正挑戰別人的直覺。有時是用判斷，判斷非理由，而是用判斷影響他人直覺。例如，這事情是好的但是未講明理由，或是大家都這樣做，直接用判斷影響直覺而沒給予更多的理由。無論如何 B 的直覺被影響，進而形成他自己的判斷，他的判斷甚至可能會形成相同推理或理由，相反他也可以因此影響 A 的直覺。

這部分是為了要將前面所說搜集道德判斷的部分詳細話，如果回到螢幕前的人所做應該選擇何項的判斷，螢幕背後的人做道德判斷可能是很複雜的事情，故道德資料庫蒐集到的為何，是蒐集到推理、道德判斷還是直覺。道德判斷、直覺、推理是否可以被更改、更正。如為可能，蒐集到的資料庫也許並沒有作用，因為在被教育後，直覺可以被更動，如果根據那些道德資料庫，以這些資料庫作為基礎進一步發展道德理由或是讓 AI 遵循，我認為太危險、太快。背後要搜集人的何種資料？如果從 Haidt 人類的道德推理來看，這相當複雜。我對於搜集資料庫這部分，抱持較懷疑的態度。或是應更精緻化，故他搜集的是人類直覺的判斷還是理性的判斷，理性的判斷在前述可以看到，理由後援有時並非真的理由，如剛剛所提及的測試一及測師二，理由只是特定、而後才加入，故如要搜集人類的理由，理性的判斷到底是辯解還是只為支持直覺。這部分的理由，我們必須小心，理由是一種有客觀性，還是純粹支持直覺，道德直覺有多樣性與衝突。

現在大家想建構所謂可說明性的 AI，可說明性所指為何，說明出的部分只是單純將推理過程講出，但並不代表具有證成性。我們必須小心的事可說明性 AI 建構出，並不代表事情被解決，可能只是將黑盒子情況講出，告訴我們的可能只是緣由、codes，仍然不是理由。或是做另一組對比，人類在做某些行動是，我們可以詢問，有些可能是 motivative reason，有些是 normative reason。如為何打人，行為者說是因為心情不好，這是 motivative reason，是動機是的理由促使他去做某些行為，這並不能說是 normative reason 去正當化行為者的行為，故可說明性的 AI 做出後，可能只停留在緣由部分或是 motivative reason。

問題與討論（敬稱省略）

吳建昌：

如果大部分的人都無法講出支持或反對的理由，AI 的限制也是與一般人限制一樣，假設找一位我們認為的道德大師，請他看上千億種案例，做出結論，AI 模擬與他看到的情況一樣，這個世界的樣態可能就是這麼多，AI 做的與道德大師也都相同，道德大師未說明理由，但 AI 做出的決定與他一致即可，這樣

會不會有問題？

蔡政宏：

除了在研究人工智慧，也在研究人類的實際智慧，有人利用列舉範例方式，談論何為「智慧」。他列舉範例，如智者達賴喇嘛、孔子，現在發現範例本身有無共通性，甚至做道德判斷時有無受所處文化的影響，如孔子與達賴喇嘛所做的判斷是否一致。這要挑戰的是，我們真有可能這樣的範例嗎？所謂的道德大師，也許我們可以先舉幾個例子？

吳建昌：

我所期待的並非普遍性的道德，這牽涉到後設倫理學的問題。即到底有無普遍性的道德，在各處都是正確的，如果不將野心放得如此高，在某種生活場域中，至少在該場域大家認為他是大師，AI 就不要出去該場域做決定，生活在這種場域中我們都認為他是道德大師。

蔡政宏：

這個想法也許會將該場域放更低，就目前情況可能是要將該場域越限制越小。如機器有所謂防呆裝置，是為了要保護人類，從這些所謂防呆裝置擴大，不管是人或是 AI 做的道德判斷，到最後越來越複雜的情境。我認為這是較接近想像的情況，如果真的有，我的回答可能是沒有所謂道德大師這種存在，我們也不完全只是透過直覺思考，我所採取的可能是 due process，在心理學或是快思慢想所使用的心靈模型。即我們有兩套運作機制，一個是直覺，一個是推理。Haidt 有個問題是他只討論 system 1，直覺的系統，較少談 reason、analytic 的系統，對於直覺系統可能有些可以做更正的地方，這是我剛剛可能的回答。剛剛問題，有部分沒有掌握得很清楚，假定有道德大師，我們學道德大師的判斷，將道德大師的判斷教 AI 嗎？

吳建昌：

AI 學習做出的決定都會跟道德大師一樣，因為道德大師所做的決定我們都認為是正確的，假設是這樣，有沒有辦法模擬成這種情境？如果要批判，你做的決定都是對的，但是中間過程缺少一個東西，沒有那個東西就無法繼續。

蔡政宏：

我有兩個回應，第一個是所謂的道德大師或是 practical wisdom，一般給人的印象是這個人具有實踐制能夠下判斷，至於理由好像是沒有，似乎在某個情境下就可下判斷，這是一種解釋，現在比較多的哲學家要挑戰這個看法，即作出該判斷完全沒有合理的理由嗎？理由到底為何？故需要將理由找出。第二個是我們要界定道德大師時，我們在判斷一個東西是否為可靠機制，必須經過長時間並不斷測試，才能賦予某種特徵。當我們認為他是道德大師時，我們必須長時間觀察他所做各式各樣各異的道德判斷，而這種道德判斷都是為我們所接受的，才能歸因於某種屬性或特徵給這個人或是機器。然而，第一個問題是，是否有個 agent 經長時間被我們觀察，而總是能做出可靠的判斷？第二個是，我們在談及道德大師時都是停留在較古老時代，當代較少找到，有沒有所謂道德大師或是智者都是相當程度被美化的可能性，也就是長時間的測試下並非所有判斷都是對的，而是他做的某些判斷，我們將他的判斷美化。這樣會有危險，我們要歸因某種屬性給一個對象是，不能只根據這個事件的判斷(single judgement) 做得很好，即認為這個人具有好的傾向，即總是能夠做出好的判斷。必須經由不同判斷個別長期觀察下，我們都認為是好的判斷，我們才能歸因於實質案例，因為過往的判斷都做得很好，所以在下次的決定也能做得很好。從此情況，我會想要詢問，也許我們應該要給一些實質案例，道德大師是我們很相信的，我們就讓 AI 向他學習，也許能學到某些東西，但我認為可能性很低，因為是想像的情境。

吳建昌：

我把他當假設，你是挑戰假設。我之所以假設是因為，假設真的有可能這種 AI 我們會去如何批判？

蔡政宏：

如果真的有我們不會批判什麼。

劉靜怡：

假設倫理案例資料庫，我當然是有些疑慮的，但如果真的有辦法做到那個樣子，AI 也進行學習，學習後會有什麼問題？

李建良：

這些問題不是都沒有答案嗎？

劉靜怡：

故我們嘗試探索，就自駕車的案例，自駕車是一個有趣的問題，理論上有很多問題是沒有答案的，謝世民老師嘗試要創造出答案，再使 AI 學習，假設是這樣。這就與剛剛建昌所說的有點類似，人自己本身也無法判斷對錯、是否倫理，我們就從頭開始教 AI 什麼是倫理的。

吳建昌：

如同父母教導小孩，父母教得好不好，其實我們永遠都不知道。他們經常都不是道德大師，所以如果真的有道德大師訓練 AI 學習也沒有關。回到現況，因為不可能有道德大師，父母同時也可能非完全正確的在教導小孩，我們都認為是可以的。那 AI 我們本來就不會期待他是道德大師，所以有些決定做錯，好像與人也會做錯一樣。

劉靜怡：

但是我們會期待 AI 在做習得的倫理判斷，還要做某種程度的理由嗎？

吳全峰：

除了理由之外，他不需要 reflection，就人而言有所謂道德大師存在的話，假設換了人的社會，會希望有個道德大師會跟我們說所有答案，我們好像也不會期待這種情況。

吳建昌：

我們不期待，可是在討論時常常會討論判斷是否夠好，好像會假設有個理想答案在。所以我們才會說趨近該理想答案是比較好的抉擇，因為在討論時，這種影子好像一直都藏在某處沒有去除。因為這樣去想像 AI 做決定，好像不夠好，因為有理想答案在那。

吳全峰：

我想問的下個問題是，假設一個道德大師存在是不可能的，社會上也不會期待有道德大師引導我們的行為時，變成我們的行為如同剛剛所述，可能是由父母所教導的，父母教導後我們會有行為存在，會在他的價值體系上。惟當我到一個程度時，會開始有反思，去挑戰他所教導給我們的東西。在這情況之下，假

設在 AI 之下，所有東西都是由該道德集給予，這個 reflection 有可能出現嗎？

吳建昌：

還是有可能。因為一開始是監督學習，接下來是非監督式學習，所以會變動，他會變成好人或更好，壞人或更壞。

吳全峰：

所以我想問的是 reflection 發生會是要件嗎？

吳建昌：

人的行為模式稱為 reflection，但內在的運作為何，其實不知道。變成 reflection 是在可以觀察到的表面，事實上在後面，如機械化部分，如何操作得出，其實自己也無法解釋。

劉靜怡：

如果連那個 AI 為何都不知道，我們要求 reflection 差別在哪？而且為何要要求？

林建中：

我猜整個的論述可能是，希望創造一個不只是模仿現有的道德，而是發現或是完成我們無法完成的道德判斷的 AI。

蔡政宏：

我所持的較懷疑論的立場，是對正要做這件事的人，不管是某些資料庫，或是將資料庫侷限在 wise 或 wisdom 不會有錯的這部分讓他學習，我想要凸顯人是很複雜，在此方式下讓 AI 學習到的東西，人的價值觀很不同。

林建中：

假如目標只是這樣，不見得會有這麼困難。他只要複製人的反應，或是多數人的反應即可，不一定要有道德大師，可以有個最大的資料庫。譬如 80%的人覺得這個不應該，也不需要判斷優劣，只要遵循大部分人的判斷即可。

蔡政宏：

這如同我剛剛提及的兩個情況，第一個是直覺先行兄妹的情況，大部分人覺得不對，但有些人不認為不對。

林建中：

那還是回到有無優劣的問題，假如該 AI 的功能是要談到道德性，不會做出與一般人反應相差太多的，只要靠資料的反射即可。這問題必須有意義就是想要做出不一樣的東西，他想要告訴我們多數人是錯的，或是某種先知，或是比所有人隨機抽樣的判斷更卓越，是否有此預設目標？

蔡政宏：

這可能更長遠，他目前目標非他來教我們，而是符合人的道德判斷，現在問題是將人放成個別的人。

林建中：

問題就回到建良老師所說的，我們沒辦法 100%確認前述兄妹的案例是對或錯，故假如我們無法確認對錯，一個叫鄉愿的方法是那就都可以，因為現在也是。

蔡政宏：

如果有人會挑戰兄妹案例，我會去幫他辯護沒有道德上的錯誤，講出去會受譴責，但這是在人類社會下會影響。

林建中：

照你的目標，我們希望 AI 做出最精準的道德判斷，即便是少數說。為何需要這樣？

蔡政宏：

可以分層次，一個是簡單的 AI 道德者，是不要傷害人類，至於本身是否為 agent 不重要，只是與人類社會互動時不要妨礙的人類的 well-being，這是一個層次人工道德，還是當作物。另一個層次是希望到高點，如剛剛提及的醫療機器人，特別是到高的等級，他們預估是可以完全自主化，這個情況下的 AI 道德判斷，

如果能比人更好當然更好，但目前做不到。初步方向問題如何教給他道德，我現在提的是教給 AI 時，不管是由上而下或由下而上會遇到的困難，我認為工程師要更細緻、小心。如果要說到更長遠的，較樂觀主義者甚至認為會有 super intelligence，現在是人造 AI，之後會推出 AI 製造出 AI，最後會有 super AI 的出現，如果是這樣，當然就超越人的智能。回到目的性，有些人正在做出機器道德，但有些時候沒有意識到背後相關的人文社會議題，我們是提醒不要太樂觀，如果 AI 是要模仿人類道德，人類的道德如何證成，道德判斷如何形成。這部分，我看的 AI 的書都未談及，他們都十分樂觀。另外是提到多數人，人數最多的就是中國，那他們所形成的道德判斷，我們要接受嗎？可能不是。所以透過案例庫的結果，我們應該要接受嗎？搜集這些資料庫應做很多資料的反省，這些資料到底要為何種用途？

劉靜怡：

你提出的這些挑戰，我認為也都是有道理的挑戰。世明老師在建立、應用資料庫的過程中，假設是 bug，整個計畫中是否有 debug，似乎沒有提到這部分，我想整個計畫中應該有 debug 的機制，我認為這些就是 bug。

蔡政宏：

我是針對現在哲學界有一套實驗哲學的做法，實驗哲學這二十年來挑戰過去哲學的做法。以前我們分析知識的做法，實驗哲學是搜集很多大學生、路人怎麼看到知識的概念，我對於整個方法不能理解。專家形成的直覺與一般人形成的直覺不同，在談知識時會意識到，對於某些特徵的敏感性會特別強調，但一般測試學生知道與否，他所使用的「知道」，我知道某人，他也會當作知道。但我們在討論「知道」時會區分是命題知識、事件知識或對象知識等。一般人不會意識到這個問題，也不知道這三類知識的目標為何，去搜集學生、一般人對於知識的概念到底能幫助我們對知識的理解有多少進展，我十分懷疑。有些哲學家反對，提出某些想法。可以看到大部分人的看法，但我認能不能這麼快就基於這些資料做出道德原則。

吳全峰：

假設資料庫成型，最後會形成標準、統一道德嗎？

劉靜怡：

現在好像是期待這樣。

吳全峰：

假設有道德大師就完全以該道德大師為主，所有東西都回歸到資料庫的檔案，到最後會變成世界上所有 AI 遵循單一的道德準則。

蔡政宏：

道德判斷會有比率問題，如果要用 80%的道德判斷要如何說服 20%的人，這些人可能會抗議，為何少數人就是代表錯的。

吳全峰：

如何去解這樣的問題？

蔡政宏：

這我就不是很了解。

李建良：

這有個方法論弔詭性，這些問題不會是從 AI 存在才發生的，在 AI 之前人類已經持續幾千年，為何在 AI 出現後，突然要為 AI 處理這些問題。

蔡政宏：

人造物有很多，但是 AI 是最特別的人造物，特別是將來會具有自主性，有自主性就毋庸透過人，可以直接與人互動。故較擔心的具有自主性的物的互動，他的 body 可能比我們優越，我們要如何互動。我認為我們對 AI 犯錯的容忍性低，對人反而較高。

吳建昌：

所以我才假設如果他是道德大師訓練出的 AI，是不是不太會反對，其實人本來也做不到，所以這假設是不成立的。人可以做大量不同意見的表達我們都可以容忍，但是 AI 好像一定要有道德大師的程度，我們才會放心，我剛剛所問的問題是要呈現這一點。

蔡政宏：

我們是否要對 AI 容忍度高一點？

劉靜怡：

是否人創造 AI 是因為自己太不完美，所以希望創造出比我們完美的 agent，更有效率做整個決策，犯更少的錯，所以才會有希望 AI 變成道德大師的期待。

蔡政宏：

如果超級智能(super intelligence)是可能的，他們預估是他會是邪惡的，人類所能做的，可能是發展較完美的 AI，這是較玄想式的，這是可能的理由。

肆、AI 機器人「道德推理」：人工智慧

一、價值問題&福祉職學

有個學者提出，在談到人類道德時，常常想的道德法則等，其實人類在做道德判斷時，會有更廣的背景，關心的是如何過更好的生活。在這部分會討論為何要做道德行為，有的是工具理由，即做這件好事會幫助達到其他目的。有的是非工具理由，認為做好事不是為了其他目的，單純做好事是人性。有人將之整理出，上面是不同對於美好人生的哲學立場，一個是享樂主義，我們活在世上是為了取得 pleasure。有人認為使得生命更滿足，有人認為是所謂美好人生就是所有慾望被滿足。另外一個是較亞里斯多德式的想法，身為人人性被滿足就是一個美好人生。故關於美好人生這裏有好幾種看法，針對這美好人生，如果詢問為何要做有道德的事，美好人生可能與道德不同。我們可以看到社會上有許多人可以過著美好人生，但是他們不見得有道德。在此如果針對採取不同價值觀的人為何，享樂主義者會認為因為他能引起你的愉悅就可以做道德的事情，所以要不要做道德的事情是條件式的。有的人會認為如果能引起生命、慾望滿足就可以做道德的事情，有人認為如果是你的天性就可以做道德的事情。

這裡要談的是，道德問題現在給 AI 道德推理，事實上如果要的是理想化的 AI 或是最終希望 AI 有人類道德的話，應需視人類思考道德問題時通常會想到生命或人生。在電影「機器公敵」中，有位警探就在批評該機器人，警探有次與小女孩搭車，整部車衝向海中，有個機器人看到變去營救掉到海中的警探與小孩，該機器人看到警探和小孩開始估算活的機率，警探的存活機率有 45%，小孩是 11%，故機器人救了警探。警探對這件事十分耿耿於懷，整部戲就是他對於機器人所包持的各種懷疑。在這提到機器人所做的就是邏輯選擇，只視某個人的存活率為何，戲中警探就提到，他認為該機器人是沒有心的。我

用較具象化的方式描述，我們在考慮道德判斷時，也許人類考慮的是更廣泛的，故將人工道德擴張談到 well-being，我們就可以來討論人工智慧。有幾篇文章在討論框架，有可能將 well-being 的概念，不只是單純說道德法則。

問題與討論（敬稱省略）

吳全峰：

所以到最後 Well-being 最後是對人類的而言的 well-being，還是最後必須是對 AI 的 well-being？

蔡政宏：

這又是另一個問題，當然對人而言。但我們要詢問的事，一個人類智者在談論 Well-being 時，有個特點是，不只考慮到他的 Well-being，甚至他的 Well-being 是將他人的 Well-being 放在自己的 Well-being 中考量，這是智者與非智者非常不同之處。如果要設計出此 Artificial 的 Well-being 時，不會如此單純只考慮到 AI 機器人本身的 Well-being，他會將人類的 Well-being 納入考量。

吳全峰：

現在設計是為將 AI 的 Well-being 考慮進，而是單純考慮人類的 Well-being。

蔡政宏：

故這是較人類中間主義，較不道德的做法。

李建良：

機器人如果講東西損壞可否罵他？是人類 Well-being 還是 Artificial Well-being?

蔡政宏：

這個問題最後還是想到人類，如果習慣性的對有人性的對象做出這樣的行為，對於做這件事的人，培養出不好習慣。

李建良：

這個理論與動物保護有一派理論類似，為何需保護動物，因為殘害動物人性會變殘忍。另一派是動物為本位的考量，會有這樣的論辯。

林建中：

物的長期存在也取得說理性？

蔡政宏：

現在又引導到一個重要但又有點複雜的議題，即有無可能有 Artificial life？我們對於「life」又有不同界定，這可能會牽涉到 Artificial consciousness 的問題。

【續上報告】

（一）Hedonism：享樂主義

第一套理論為享樂主義，享樂主義認為 pleasure 構成 Well-being，哈佛大學哲學家透過設計經驗機器，因為家庭美好的經驗與是否為家庭美滿為兩件事，可以有家庭美滿的主觀經驗，但事實上可能家庭不美滿。一個是事實客觀層面，一個是經驗。如果針對這個說法，他設計了一個經驗機器的實驗，這個實驗很多會用到駭客任務的版本，牛津大學有出書針對駭客任務這三部電影引起廣泛的哲學問題，他們將問題具象化、更生動，其中一個在談論經驗機器。如果在母體世界中，他認為自己有頭髮也與人類在一起，事實上只是在培養皿中，培養皿中許多電線刺激大腦，現在我們所看到的光線與物體等，不過是接收反射最後在神經系統產生作用。透過這樣的原理講述，事實上可能沒有手、腳，但刺激相關部位你會認為自己有手、有腳。在經驗機器中可以灌輸、設計，可以先想好自己想要怎樣的人生，就又進到經驗機器中，就可以有這樣的人生。他透過此詢問，是否願意按啟動鈕？很多人不願意，這表示經驗可能不是我們在 Well-being 的唯一考量，經驗與愉悅可能很重要，但非唯一考量。這套理論的缺點是將兩個混淆了，事實上他想要敘述的是慾望，我可以有美好家庭的經驗，跟我有個美好家庭的慾望，這兩者不同。慾望滿足是事實上滿足，不只是有經驗。

（二）Desire-satisfaction theories：慾望滿足理論

第二套理論為慾望滿足理論，何謂美好人生即慾望被滿足，慾望滿足構成你的 Well-being。例如，某求在台積電工作，這時對比剛剛的理論，他只是想要有台積電的工作，作夢也可擁有此主觀經驗，但是慾望沒被滿足。必須事實上在台積電工作才有滿足慾望，故兩者為不同層次。但同樣也遇到一些問題，慾望滿足並沒有美好人生，例如吸毒。

（三）客觀表列理論

什麼是美好人生，先列出一張表，表上都是客觀上好的東西，這些獲得就可以使你獲得幸福。像是哲學家認為的幸福可能與一般人所認為的不同，如意思能力自由，有的會認為生命、知識、美感、經驗都是，不同的哲學家會有不同的客觀表列。但問題在於，有些情況下在客觀表列上，但這個人並不幸福。例如客觀表列上有生命，可以設想之前安樂死的主播傅達仁，他擁有生命，但是他認為生命並帶給他幸福，所以他尋求死亡解除痛苦，故生命是否為表列上的一個項目呢。有時有人認為幸福但不在表列上，例如亞里斯多德認為能沉思的人生是最幸福的，這是他的想法，有些人的確是如此，但表列上並未列出。可以思考，有人會將生兒育女放在表上，但有些人認為並非如此，OUP 這本書論證生小孩是個不道德的事，裡面有些論證，當然也是十分爭議。到底表上應應放什麼東西，為何放上的東西就是好的這也必須被考量。綜上而言，背後還是有點依靠直覺，一般分成主觀主義與客觀主義。

主觀主義即何謂 Well-being，完全由自己的主觀態度決定。有人認為主觀主義是不對的，因為個人所需求的並不一樣會使自己過得更好，這是一個反主觀主義直覺，當然也有許多案例。另一個是客觀主義，有些 Well-being 並非由態度滿足，特別是父母親會認為某些東西是對孩子好，不管孩子需要與否。反客觀主義者認為，小孩通常會認為你認為好的我並非也是同樣想法，對我而言與我的生命沒有任何關聯，這是反客觀主義的直覺。在反客觀主義及反主觀主義都有相關直覺支持者存在，如果要將人工智慧要把某套 Well-being 給 AI 學習，在這會有衝突。

伍、結論

建構 AI 道德不論是由上而下或由下而上都會遇到某些問題，最主要是由下而上是想要 AI 學什麼。如果要建構人工智慧同樣也會遇到某些直覺上的挑戰，是反客觀是對的還是反主觀是對的。

問題與討論（敬稱省略）

李建良：

剛剛提到機器人的三法則，Frank Pasquale 新書發表，書中一開始就從三法則講起，他提出四法則，基本上是認為三法則太簡單，剛剛你有論證這些部分。

劉靜怡：

該四法則是因為於 2017 年還是 2018 年在 Ohio 州的法學院的會議上，那天的

主要討論主講人是 Jack Balkin，雖然是憲法學者，但是做很多法理學的研究。他使用機器人三法則，Frank Pasquale 是他的學生，他們就找他當對談人，他的四法則是那天的對談稿。

李建良：

謝世民老師有個想法，希望教 AI 讀康德的作品。我最近在寫文章就有看康德的作品，康德最重要的思想是純粹理性批判，而我要談的是，基本上我是閱讀原文，後來發現英文與中文翻譯有很多錯誤，故是否要讓 AI 學習德文版本，就會有基本性方法論的質疑。

吳建昌：

我的聯想是，在醫學上，如果已經有科技可以運用假設有疾病可以治療，那技術已經可以讓病患治療到比一般人好，有些人會認為為何不讓他再更好，為何一定要恢復到原本情況。所以如果真的要發展所謂道德的 AI 或是 AI 可以做道德判斷、推理，的確這樣的想法可能會再出現，即為何我們不將 AI 做的比一般人再更好，基於這樣的期待，可以理解因為我們認為做不到，所以會採取較懷疑的態度。我有種直覺是，會期待這種 AI 做出的東西會是沒有地域性、局限性，至少要有好到這種程度我們才會放心，惟事實上是做不到的。我們還期待會有 reflection、reasoning，甚至有我們都無法找出的東西，最好 AI 都能偵測出。我們才會放心讓這樣的人造物，充斥在我們世界中，惟如果不是人造物，我們是人，人一直在犯錯，可是我們還可以容忍，而人造物是不能如同人般被容忍的，因為就算跟我們一樣還是不夠格生活、充斥在我們的世界中，扮演這樣的角色，必需好到何種程度。如果我們都無法預期能夠好到何種程度，也不知道應該期待 AI 要多好，我們還是小心為上，這是我今天的感受。

蔡政宏：

有些較樂觀的人會說的很美好，以細節上觀之，我認為還是有很多問題，有兩種可能的問題。一個是他是被新聞媒體等誘導，另一個是 AI 設計者真的不知，如我們在討論哲學或是道德問題，對於其中要注意的細節，並未強調。包括協助心理學的審查，較涉及道德討論，我們看那些實驗，其實都未討論何為道德，對於道德界定很簡單，小孩子有無將東西給別人叫道德，以此做了很多實驗，我都抱持的懷疑的心態。一部份是讓心理學、資訊科學不要做這些實驗，而是提點工作，若他們能做到注意到當然更好，我較接近您的看法，即這部分我還是有很多存疑，現在在說 AI 必須做到反事實的思考，在哲學中關於這些反事實，也是有很多討論，我不知道他們運用到何種地步，我認為還有很多層級。

可以期待高等級，惟很多討論必須建立在經驗上有何，如果沒有，都是較想像式的看法。

吳建昌：

剛剛提到的人工法官，如果我們期待能夠放心將事情交給與有道德判斷能力的AI處理，想見還有很多事要做，我想法官權力更大，期待恐怕會更高。

林建中：

假如我今天想要生產具有高超道德能力的 AI，以下有三個選項要訓練，不知道會建議哪一種選項。第一個是回到剛剛所講的，擁有最多的資料庫，盡可能的將資料庫分類，譬如這邊是中國、亞洲、美國的，盡可能搜集很多人的道德判斷，讓他能夠反映一般的判斷。第二個是給一群相對道德較好的人，挑選過的範例，如剛剛所述的道德大師，讓 AI 學習。第三個是我們不使 AI 與人學習，我們讓他自己運作模擬，有時我們其實不知道為何會得到這樣的道德判斷，譬如兄妹性行為可能想到的是某些原因，其實我很懷疑，以前人認為兄妹間不能發生性行為是因為基因問題嗎？還是有不同考慮在其中。我不知道道德的判斷從何而來，給他模擬案例讓他自己運作，計算可能的結果，幾個運作下來，看看能夠觀察到什麼。可能與人類學習模式較不同，或是相同的。

蔡政宏：

第三個讓 AI 自己運作，結果好壞如何決定？

林建中：

他自己判斷，因為我們也無從得知。如古代人近親通婚會認為是血統純正，後來才發現不太對勁，為何智能障礙的小孩比例較高，故是後來才發現的，但是很長一段時間被認為是道德正確的。我的理解是這樣，所以你的判斷是取決於將時間點切在哪裡嗎，假如切在 18 世紀，說不定那時還是覺得兄妹發生關係是正常的。所以我們只是用自己的能力侷限，他與人類學習，或是跟我們現在最好的狀態學習，或是最多數的狀態學習。第一組是多數，即最符合大眾狀態，是我們認為可被選擇的最好的結果。第二組是定邊界讓 AI 自己運作，讓 AI 告訴我們什麼是好的，因為如 2、3 世紀所選出的價值，可能與現在的價值不一樣，我們無法預估二、三個世紀後的基本權為何。故您認為哪一種是道德的追求方式。

蔡政宏：

第三種與機器公敵的情況很類似，後果之一就是 AI 為了保護人類，將人類全部關在家裡，因為他發現人類的黑暗面等，最好是不要讓人類出去，那部戲的結果是將大家關在家。人類當然會不贊成，但是裡面有個機器人名為 Vicky，他的推導是很邏輯的，他認為放任人類互動，其實問題會越來月多，最後會導向人類滅亡，這是一種第三種情況可能的結果。

林建中：

那只是一個可能，說不定會找到讓人類更開心，對於地球、動物也更好的方式，幫我們重新創造出新的道德標準，應如何看待動物、植物、礦物。

蔡政宏：

當然有其他可能性，但要回到一個問題，什麼是道德？另一個問題是人類要道德做什麼？道德一個很簡單的講法是，我們最好的狀況是人類在一起生活，在一起生活時根據某些彼此可以預料或是接受的規範作為互動，以此獲得較好的 Well-being，道德不過是這樣，不見得是道德會讓人更高超等，其實很單純是讓群體中的私人的 Well-being 彼此可以較高。在這樣的看法下，道德應回到人類自己做，因為機器沒有我們的身體、需求、慾望，不了解人類需求、慾望、價值觀為何。如果道德目的是讓大家 Well-being 更高，讓大家活得更好，我們將此權利、動機交給機器人，但機器人少了找出這些規範的要素，因為他沒有我們的慾望、需求、身體，這樣建構出的道德，即使人類作出反抗也無法理解，例如人類與父母、小孩間的關係與他人父母、小孩的關係都不一樣，這個 AI 如何理解，故讓他自己運作會有很大問題。第二個資料庫，當然較接受第二種具有好的資料庫的情況，專家，什麼叫做專家，其實要去學習的是掌握模範，在掌握很多規範後，之後類似情況可以做出認證該規範，可以於毋庸思慮的狀況作出反應。我可以同意找出所謂的道德大師等，將他們做出道德判斷的規範認證的方式做出來。問題是該規範如何界定，簡單的還可以，但通常我們面臨到的都是複雜情境，大部分都是討論道德兩難等較複雜的情況。故所謂道德大師，之所以是大師就是在專門處理困難的道德問題，故我們必須給與夠多的這種困難情況，才足以成為大師，而非只是有一個道德高度。難的是在道德抉擇做出所有人的 Well-being 都兼顧到，這時才會稱其為道德大師。故要如何界定，要給他很多的困難的情境，如同我們如何判斷一個劍擊高手，非僅擊中一次，而是一次兩次不斷至百次等，我們才會認定他為劍擊高手，道德大師也是。故到底有無這樣的道德大師，我抱持懷疑態度，如有當然很好。回到你的問題，如果是我選擇第二類較好，第三類我不放心，第一類的資料沒有消化、整理過。

林建中：

惟第一類就是目前國民法官所要採用的，我們希望相信專家又不相信專家，我認為機器的好處是與我們的運算速度不同，可以不受人類運算速度的限制，應可突破思考模式，故我自己是較傾向選擇第三個，因為第一個和第二個無非是重複人類的成就。

蔡政宏：

我不清楚剛剛所提及的國民法官，我的想法是一個法官要做決定時，他所依據的法條、生活經驗，但他面對事情的複雜度、不同的人，有時超出其生活經驗，所以那些所謂國民法官所提供的不同的生活經驗。

林建中：

同樣的論理也可以運用在剛剛的情況，因為道德大師可能只是某一限定的道德難題的大師，可是一般人所碰到的可能是簡單的，如是否要超車、按喇叭等。

蔡政宏：

我們對於道德難題的界定可能不一樣，超車為何是難題。

林建中：

譬如前面開的很差，或是在玩手機，基於禮貌我們可能不希望成為一個很激進的人。

何之行：

只是自己認為會不會影響到別人，但是對於被影響的人而言，其實如同剛剛提到的例子是完全不同。

蔡政宏：

我們討論的兩難，即 Well-being 的影響程度會很大，如同我們提到火車難題，Well-being 如果可以都做一個物品的話，或是有很多慾望構成 Well-being，有些慾望有中性地位，這個慾望不被滿足或是會受到影響，其他的慾望都會受到影

響。例如想要存活，生活或生命這個慾望，會影響到其他的慾望，但可能不會影響到中心慾望。慾望之間具有某種結構，我們提到道德兩難，通常是涉及到那個人的 Well-being，慾望或是某件事被影響，他的 Well-being 動盪會較大。故剛剛超車的問題，造成的道德難題對我而言較小。

林建中：

我在思考，第一組的對比是，我們應該從更多人還是更少人身上認定何謂智慧。譬如，很多人決定要吃什麼是看美食家的部落格，有的人是看 google 評論，會認為評論很多就相信，這只是個例子，我相信兩種方法都有在操作，有時相信專家，有時相信多數。故目標為道德上的精進的話，到底哪個系統會被推薦是我想問的問題，因為顯然現在假如要做資料庫，或是被訓練，應先找到訓練對象或方法。

蔡政宏：

我較站在道德問題放在 Well-being 角度觀之，Well-being 我採取的是接近慾望滿足理論，但該理論我採取的是變形，稱為態度成功，但這部分有些複雜還在建構中。如果回到較單純版本就是慾望滿足理論，慾望滿足理論還會遇到很多問題，慾望滿足理論可以化解較多問題，也可以較多的尊重他人，如台灣社會中有很多爭議問題，如同性戀、墮胎問題這些常在應用倫理學會討論到的，某些人的價值觀涉及慾望，如以單一的客觀表列，會有一部分的人反對，即這些人所想要做的、價值判斷，會在客觀表列會被當作是不對的。如果現在是 AI 要做這件事，這個 AI 會變成很複雜的 AI 體系，他要去關照一個人幸福時，必須知道慾望有哪些，哪些是他的核心慾望，必須針對此人哪些慾望可滿足、哪些不行，才可以關照到這個人的慾望，慾望滿足的極大化就是幸福指數較高。如果 AI 要學習並進入人類社會，我認為應該要有這樣的複雜度，資料資訊處理系統應要更好，可以發展如何偵測到人的慾望，例如機械公敵中需要偵測到警察的慾望，即小孩的存活比他自己的存活更有價值，必須協助警察滿足他的慾望，惟想像中的想法。

何之行：

他所追求的不是確定的東西，而是希望有個空間可以同意。

蔡政宏：

只是這部分我還不清楚，他的偵測系統辨識人形，從人形的喜悅就是笑，但是

人人的真實情感如何偵測，較不理解如何偵測的，也許可以從血壓脈搏等，只是技術上我不清楚。

李建良：

兩個問題，一個是法律人有何能耐認定道德為何？哲學屆或是所謂的倫理學者，是否同樣有能力說何為對錯，還是提供一個理論框架。第二個是剛剛建中談到的第三個類型，讓 AI 自己發展，有個前提是他有能力發展，所謂變成有強 AI 的傾向，才有可能自己發展。故於此有個前提是，將來的 AI 是一個 AI 還是多個 AI。AI 如果可以自己發展，理論上他們也會產生多元價值，我們好像一直假設他只有以單數概念看這個問題，他有可能是複數或是之間會產生衝突，提供大家思考的問題。

第四次會議紀錄	
時間	109 年 10 月 23 日（星期五）
主題	開放銀行與消費者賦權的想像與挑戰
講者	臧正運（國立政治大學法律學系助理教授）
內容摘要	

壹、虛擬與開放的金融體系質變

一、金融體系的本質是記憶與信任

先行回顧過去幾年間金融市場發生了何種質變，乃至於現金金融科技會需要從另一視角看到。很多金融理論、法律制度、規範層面的創新需要發生。將之稱為虛擬與開放的金融體系質變，金融體系的本質是記憶與信任，在一個金融體系中有許多陌生的交易主體，其間之所以能締結交易、產生金流，再下一道交易，背後有個很重要的基礎是有一個機構，在陌生交易主體間擔任信任攻擊者的角色。該機構就是我們所熟知的金融中介機構，也可以用銀行理解。故銀行所扮演的角色是，接受政府高密度的金融監管，換取作為市場信任攻擊者的對價，扮演擔任市場攻擊角色。為何要市場攻擊、信任攻擊？因為沒有信任，交易就無法發生，故這是金融中介機構存在的第一個要務。隨著此要務的建立，開始在金融中介機構中慢慢積累出，我將之稱之為以金融機構主導的記憶體系。可以想像我們將錢存在銀行，投資理財透過投信投顧業者，買股票透過經紀商，買保險透過人身保險公司、財產保險公司，這些金融機構扮演的角色是為我們的交易提供承載的記憶體系。故以此簡化版方式理解金融體系的本質可是說，金融體系長久以來在做信任的攻擊，以及記憶的維繫，信任的攻擊與記憶的維繫不是只有金融中介機構可以做，但是過去都是以金融中介機構為主導

而運作，隨著資訊科技的興起與發展，這些根本角色可能產生改變。故下文要談的是，在此質變中如何看待開放銀行發展的脈絡。我將此兩個重要的需動力一個稱之為虛擬化、一個稱之為開放化。

二、虛擬化

如果以前信任的供給、建立與維持是建立在人與人之間實體的互動，現在信任建立、供給與維繫產生轉變，這個轉變已經發生好幾十年。如以前銀行中看到實體行員，現在可能透過電話、網銀的 App。可以看到信任這件事被建立了，維繫需要一些方式幫助。會看到虛擬的通路試圖營造、維持信任，這件事在信用科技的時代被推到了極致，故於年底三間純網路銀行即將式開業，這些純網路銀行有個特色是，在純網路銀行開戶就會透過網路，故行員無法實體面對面進行 KYC（know your consumer），需進行線上 EKYC 的方式，必須有虛擬的方式維繫與客戶間的信任關係。這些事想像容易，但運作上有難度，這是虛擬化對金融體系帶來的第一個挑戰。

第二個挑戰是，除了看到從實體通路轉虛擬通路的金融服務與創新之外，同時也看到金融體系過去承載金融體系足以運作的載體或載具，將之稱為貨幣，最基本原始的單位。此貨幣通常是法定貨幣，但現在虛擬貨幣、密碼資產興起，這些以虛擬通貨為基礎所搭建的新的記憶體系，是與傳統由中介機構主導的記憶體系截然不同。會產生實體貨幣為基礎的體系，與虛擬貨幣為基礎的記憶體系相互爭奪記憶權的過程。事實上，過去從比特幣的興起到各種不同名稱的虛擬貨幣，前陣子 Libra 指出要推行 Libra 幣，承擔類似全球央行的職能，可以看到央行的替代方式，除了 Libra 在做的事不應該存在外，央行自己也說 Libra 在做的事央行也可以改變，故全世界有 80%的央行正在研究、實驗，中央銀行所發行的虛擬貨幣或是數位貨幣。在我的角度而言是一個記憶權爭奪的過程。那些實體掌握金融貨幣體系的人認為，有一整套金融交易是不在他掌控下的記憶體下發生的。應如何看待？背後有許多監理規範的啟示，舉例而言過去幾年台灣最常討論的洗錢防制，洗錢防制是所有金融領域與財經刑法領域研究的學者朗朗上口的事項，但當虛擬貨幣興起時會看到，很多主管機關認為虛擬貨幣興起不利於洗錢防制，會帶來許多問題。有些幣不容易加密、追蹤，如門羅幣。惟從另個脈絡觀察會發現這是金融記憶權的爭戰，在此記憶權的爭戰中觀察，會看到虛實之間彼此相互整合的狀況發生。一個最近的例子是，美國的 OCC、ACC 正式頒布一個指令，以前美國銀行是否可針對穩定幣的提供者、虛擬幣的提供者，收了大量的法定貨幣，將錢存托在商業銀行中。過去這件事存在而被實務上接受，但是美國主管機關沒有公開聲明允許，在上個禮拜美國的 OCC、ACC 共同發布一個指令，聲明往後從事穩定幣，且穩定幣是在託管帳戶下運作、發行，可以把法定貨幣託管在商業銀行。故會看到未來有越來越多虛實整合的場景。前天 Paypal 聲明可以接受加密貨幣，如比特幣、以太幣，做為支付工具，將這些加密貨幣換成法定貨幣，在 Paypal 可以提供支付方案的兩千六百多個商家當中進行交易。故會看到許多虛實整合的交易場景發生，這

件事對金融監理的啟示是，金融記憶的體系開始被擴張。在此擴張與整合的過程中會產生破口，這對金融監理產生一定程度的影響。這是我觀察到的第一個重要的趨勢，故純網路銀行、虛擬通貨的崛起，大概都可以在此脈絡中被理解。

三、開放化

第二個是開放化的趨勢，這與下述談及的開放銀行有很大的關係。如果金融體系最核心的功能與基石是供給信任，信任關係是建立在我作為金融消費者或客戶，我將信任託付金融中介機構，基於對金融中介機構的信任，我相信主管機關會近期監管職責，基於這樣信任可以透過該機構進行各式各樣的交易。現在會發現這些金融機構，在無形中隨著商業需要、技術發展，會將信任延伸出，故於金融監理的研究中有一重要環節為委外作業的監理。可能今天銀行委託，可是銀行將某些需要自己處理的事，認為可以找第三方服務業者提供更便捷成本更低的服務，本質上是將客戶信任延伸出的過程。在此延伸過程中產生傳統監理的意涵，傳統監理的場景發生在銀行與客戶，當銀行將原客戶交付的信任的延伸後，那段信任的延伸與賦予應如何看待，誰可以管理此過程中所會遇到的風險，這是第二個過程。

今天要講的開放銀行與此有很大關係，人們未來可以透過某些技術方案、平台興起，要求與客戶往來的商業銀行在客戶的授權與同意下，把銀行所掌控的金融消費者資訊，如帳戶、交易資訊，轉給客戶所指定的第三方服務提供者，由此第三方服務提供者對客戶提供更多的服務，本質是一個由金融機構在將信任延伸出給第三方服務提供者的過程。此過程仍會產生信任的建立、供給維繫方式的挑戰。同樣也會造成記憶網絡的擴大，可以想像當這樣的記憶網絡被擴大成將各式各樣的第三方服務提供者涵蓋，這個網路擴大的過程會形成破口。

貳、資料賦權的動能

過去兩年再談金融機構的管理時，資安變成一個很重要的議題，原因是大數據時代下，金融機構必須要承載非常多的資訊，這些是一個記憶網絡的擴大與深化，如何確保資訊大量、快速、頻繁、多元的產生，但是記憶網絡的運作又不出問題，故挑戰了資訊管理的能力、資訊安全的能力，可以粗淺地以虛擬化與開放化主要兩個推定力量理解，金融體系的改變。我認為因為這樣改變慢慢進入一個可能性，我將該可能性稱之為資料賦權的可能性。這與劉靜怡老師所說資料驅動有點異曲同工之妙。所謂的資料賦權是隨著各種資料的運作，資料可以授權給兩者層面的人，第一個是業者，故業者透過拿到資料，故可以單件或催生各種的服務，數據某方面授權的產業的參與者。但是從另個角度觀之，數據很有可能授權的控制者本身、或是消費者本身。所以我將此概念稱之為資料賦權，會產生一些與開放銀行有關的問題。在此資料賦權與虛擬化的金融體系質變之下，背後還有技術的興起所做的支撐，賦予其源源不絕的動能。過去可以看到與科技發展有關的用語，不管是機器學習、深度學習、類神經網絡、自然語言處理、分佈式帳簿技術、雲端運作與儲存等。這些東西從我的角度觀

之都與資料密不可分，舉例而言，機器學習與人工智慧的本質是處理資料、分析資料，也許處理資料是為了進行預測，故處理分析。密碼學的目的是為了加密，生物辨識是為了幫資料做核實識別。當後面所談到開放銀行時，是透過一種公開、標準的應用程式介面，進行資料的交換與共享。再談雲端運算、儲存、管理時，其實是資料的處存與管理。故這每個技術的興起與資料有關，具備以下幾個特色是，這些技術使用的門檻比以往來得低，可以以比十年前低的技術成本使用這些技術。第二個可以使用多元的方式，同時使用上述所提的兩種以上的技術。第三個因為前面所提到的條件導致，有一種類型的業者開始出現，是否有可能因資訊科技進步，進而做金融體系傳統維持、提供的功能，也就是信任的攻擊者與記憶的維繫者，故可以如此理解。

問題與討論（敬稱省略）

邱文聰：

資料賦權的受詞是誰？

臧正運：

受詞可能是業者、消費者。

邱文聰：

因為標題是消費者，可是剛剛所說的是包含業者的。

臧正運：

是的，以下會進入正題。

參、資料與金融

資料與金融產生的關係，這樣的關係又如何影響法律的角色與監理的認知。資料有 4V (variety, velocity, volume, veracity)，此 4V 意味的是記憶體系的承載能力要很強，記憶題系的防誤能力要很強，記憶體系的韌性要很強，故對金融監理會產生一定的啟示。第二個會發現資訊流的掌控者逐漸成為記憶與信任的供給者，可以看到現在所有大的電商平台、社群媒體無一不在從事與金融相關的服務，只是在不同地域的問題，他們發現他們可以做原本金融中介機構的是，惟對於監理而言會產生某種程度上的挑戰。資料管控能力會在業者與

消費者間產生一定的消長，過去對於銀行幫我們管控的資料，唯一我們能掌控資料方法就是帳簿，帳簿可得知做了哪些交易，僅止於此，沒有其他方式更有效率掌控我放在銀行的資料，故我與銀行間產生存在某種不對稱，這種不對稱可能是資訊的不對稱、資源的不對稱、可能是資料控管能力的不對稱。隨著科技的發展慢慢開始有改變，資料的管控能力不再如此向業者這麼傾斜，對消費者而言，可能也有一定程度的提升，但是提升意味什麼？且應如何看待？故於此，從法律與監理的角度觀之有何種啟示。第一，會發現信任的維繫越來越困難，但仍非常重要。從傳統金融監理的理論觀之，信任就是金融體系穩定的根基，仍需維持，但該維持越來越困難。第二，會發現有各種不同的記憶體系出現，這每一個記憶體系之間是否相容、可互通，如何確保相容的過程中不出問題、沒破口，這是第二個重要的趨勢。第三個，有無可能有較好且可可能的責任分配機制，假設沒有公平、合理的分配機制，會使資料賦權沒有辦法發生。第四個，會需要重新省思資料在整個過程中產生的作用以及倫理問題，甚至因為使用資料的方式所產生金融分配正義的問題。舉例而言，透過機器學習、大數據分析可以給予信用評等、信用評分，進而授信給消費者，很有可能基於這些資料所做出的是錯誤結果，或是有歧視性的結果。這件事會影響到整個金融市場上的信用不公平的分配，這個金融正義的問題在整個過程中如何被解決、理解。最後要回答上述四個問題，對於金融監理機關而言擁有巨大挑戰，因為監理機關不一定具備資料管控能力、科技能力處理這樣的問題，故我的研究中主要放在監理科技的研究，以監理官的角度而言，科技可以對監理官帶來何種改變，可以為監理官做什麼。

肆、開放銀行

一、浪潮

舉個場景幫助理解開放銀行為何，開放銀行並非銀行 24 小時、一個禮拜七天都開放即稱為開放銀行。假定每位手上都有四張信用卡，在使用信用卡時常常會有人告知走到某地消費應用某家信用卡，走到某地消費應使用某家信用卡，因為在此消費綁某信用卡會帶來更多現金回饋，對我而言，假設我擁有四、五張信用卡，我碰到的問題可能是我不一定清楚信用卡提供我的福利是哪些，在何種消費情境應使用何種信用卡，故幻想未來是否可能有種服務，建立在手機上，一個平台供應者就是一個記帳軟體、信用卡管理軟體，透過授權跟所有往來的信用卡發卡銀行簽約，可以在我的授權下從我的發卡銀行取得我的資料。假定這件事發生，這樣的網路平台業者可以做何事？可以透過手機提供的位置，當我走到 711 進行消費時會自動跳出，用哪張信用卡消費會得到最高的回饋。如果不滿意，可將其他後備選項叫出，仍然選擇用我堅持的信用卡消費，這個產景或可以建立在我所往來的眾多銀行帳戶下，假設與十間不同的銀行往來，每間銀行都掌握我不同的交易歷程，對我這個人在金融市場上的往來樣貌有不同程度的瞭解，有無可能有一個 App 幫我做單一介面的管理，清楚地告訴我每個月金錢的流向，從哪個銀行進出，未來要如何使用錢，投資如種基金

生品，購買怎樣的保險產品，哪些是我所需要的哪些是我所不需要的，透過一個平台滿足上述所有的需求。此必須建立在重要前提上，有一個第三方服務業者，可能是科技業者、平台業者，必須能夠與我往來的銀行或是發卡銀行要求與我有關的資料、帳戶資料，甚至是交易資料。故開放銀行是將剛剛所述的看似便利、美好的願景發生所需要的法治基礎設施。

故首先先看開放銀行的定義，開放銀行可視為開放金融體系的一環，未來會常聽到開放證券、開放保險各式各項的東西。係指銀行自發或經法規要求下，將其所管理之資訊（如產品、帳戶及交易資訊等），以開放應用程式介面（API）或其他安全之方式與其他銀行、支付機構或第三方業者（Third Party Service Providers ）分享，藉以強化金融市場之有序競合、激發創新金融服務模式、確保消費者的資料主導權，進而營造多元創新的普惠金融生態體系。

這些事基於何種目的需要被推動，第一，希望金融市場有更多元的競爭。大家對於銀行所銷售的商品沒有我們想像中的熟悉，舉例而言，如需要貸款，要如何知道哪家銀行可以提供我最合適的貸款，在現在低利率的時代不必然就是台灣銀行公教人員優利貸，可能是別的民營機構的銀行，故不一定知道。不知道的原因是因為沒有一個合理介面幫助我知道，故如上述的情況能夠發生，就可能導致市場上會有更多競爭、更多的金融服務模式。消費者在此過程中，針對被金融機構所管控的資料取得一定程度的主導權，達到普惠金融的目標。普惠金融就是讓每個人都以自己能負擔的成本，接近使用所需要的金融服務。以上是對開放銀行的定義。於此我們知道開放銀行的重點不在於銀行是否要 24 小時開放，關鍵在於，銀行需願意在客戶的授權下，把客戶控管的資料，在客戶的同意下用安全可靠的方法，分享給客戶本人或客戶所指定的第三人，也很有可能是客戶所指定的第三方服務業者，由第三方服務業者與客戶所往來的銀行索取資料，可能會發生兩種不同流向。

開放的標的為何，開放標的就是資料，惟資料有許多不同類型。以金融場景而言，資料可以區分為完全公開，與金融機構營運有關的資料。例如，在哪有分行、每天運鈔車的方向、頻率，這些資料為公開的資料。這些資料對於金融機構而言，不會有誘因不分享，可能會願意使用標準化格式分享。金融機構所提供的產品能用 API 的方式提供，對於金融機構而言沒傷害，因為就是他的產品資訊，可能只是要做比價網站而已，故產品資訊對金融機構的阻力較小。惟與客戶有關的資訊，如姓名、住址、電話、收入、動產與不動產，可以認為是個資，對於銀行而言，雖是個資但該帳戶是客戶經過某種契約關係開立的，這個記憶體系是銀行為客戶維繫的，沒有負擔任何成本嗎？而後又碰到帳戶資訊，帳戶資訊可能是在某銀行開了一個活期存款帳戶、外匯存款帳戶，開了不同的帳戶。請問要將此帳戶資訊在客戶的授權下分享給第三方服務提供者？金融機構產生懷疑，這是否為個資？如果是，可以做何種主張。銀行應有參與帳戶的維持、編輯、記憶的工作。這件事更進一步會涉及到帳戶交易的歷程，如是否可要求銀行因為客戶要換銀行，不熟悉過去往來的歷史，故要求往來歷史

提供，且是使用提供開放 API 的方式提供。銀行會認為交易資訊是你的資料嗎？銀行可能會抗辯這是你的資料也是我的資料，為何要分享。還有一種狀況是，客戶可能會要求授權在銀行帳戶中發動某種類型的交易，故在網路上消費選擇商品後，要求銀行從中扣款，且將啟動支付指令的權限給第三方服務提供者，此對銀行而言問題更大。還有一種類型的資料是被銀行加工過後的資料，銀行針對客戶過去往來的紀錄給予內部信用評級，這種資料要開放嗎？想當然爾銀行不可能會開放，因為是營業秘密。故在談開放銀行時會依照資料屬性討論開放範圍，國際上有二十幾個不同的國家正在推行開放銀行，國際上大部分認為公開資訊一定可以開放，產品資訊鼓勵開放，客戶的帳戶資訊可以開放，帳戶中的交易歷程可否開放每個國家的作法可能有點不一樣。啟動支付的交易有些國家做得很前面，如歐盟，可能是為了確保支付業者與銀行業者可以站在某種公平的競爭地位，故會認為可以推動。但是到目前為止，尚未看到要求銀行提供加工後的資料的，這是從開放的標的看開放銀行這件事。

除了開放的標的，也就是資料的類型外，還會討論到資料的種類以及資料的權限。故一般而言，給第三方業者的權限有兩種，一種稱為 Read Access，一種稱為及 Write Access。前者是很單純給予讀取資料的權限，後者是將啟動某種交易，特別是支付類型交易的權限開放。國際上，歐盟在支付領域上有提供 Write Access，澳洲只有做到 Read Access，某些地方會有不同的搭配與類型的場景，這是一般的分類。

二、目的

（一） 促進金融市場各行為主體間的競爭與互補

為何有開放銀行？在開放銀行的討論上主要有幾個先進國家，所謂先進不是真的做得很好，而是較早開始從事。如歐盟，歐盟很早討論 PSP2 時，就有在討論開放銀行，只是歐盟從來都不在法律規範中使用開放銀行這幾個字。而英國更早於 2014 年時針對零售銀行業者進行系統性調查，該調查是由 CMA 進行，是一個競爭法的主管機關，針對市場特別是零售銀行市場做廣泛調查，發現原來英國人不喜歡換銀行，且不清楚銀行所提供的產品、收費費率，被九大銀行長期宰制。他舉了個例子，要求英國人換銀行，英國有九大著名的商業銀行，該九大銀行市佔率 80%-90%，幾乎八成到九成的英國人都跟九大銀行往來。他發現叫英國人離婚都比叫英國人換銀行容易，故銀行有時收費不透明。例如透支，與銀行可能可以透支一定額度，可是透支銀行會收取手續費，英國發現每家銀行的透支手續費不一樣，原來客戶不清楚透支的手續費如何。故英國的競爭法主管機關發現此問題，認為需要改善零售銀行被九大銀行獨占、宰制的現象，因此要採取競爭法上的補救措施。故英國推行開放銀行的動機是，基於重新改變市場以及長期不正競爭的狀況，所採取糾正的措施。是從競爭法主管機關所發動的措施，所以做法是九大銀行掌握如此高的市佔率，要求九大銀行出資設立公司，稱謂 Open Banking Implementation Entity(OBIE)，由該公司建置開放 API 相關的技術規格、資安標準、監管治理的機制，目的是讓英國

銀行的客戶可以在他們的授權與要求下，讓九大銀行將為英國消費者所控管的資訊，在消費者的同意下，分享給消費者所指定的第三方機構，這是英國的脈絡，初衷是基於彌補市場上競爭不足的狀況，所推展的結果。

問題與討論（敬稱省略）

邱文聰：

九大銀行成立，又使消費者指定？

臧正運：

因為決大部分消費者都是該九大銀行的客戶，該 OBIE 的角色只是針對 API 規格標準化，故針對 API 的規格標準化後，九大銀行會依照該規格、消費者只是將資料傳輸。

邱文聰：

故會有 OBIE 以外的第三方服務業者？

臧正運：

OPIE 做幾件事，第一，測試。如 API 要上架時，OBIE 可以幫忙做測試，是否符合標準。第二是標準化，銀行與第三方業者在有開放 API 之前也會做雙向合作，但這個規格是不一致的、不標準化的，故 OBIE 的存在就是將規格一致化，使大家用同一套規格標準運作。OBIE 還提供一定程度的爭端解決，只是提供一個渠道，兩個銀行間或是兩業者間在此事上有紛爭，透過 OBIE 所搭建的平台彼此協調，不實質處理消費者的糾紛，只是做個平台將有紛爭的消費者串連。

【續上報告】

（二） 催生創新的產品與服務，提升客戶體驗與消費者福祉

（三） 活化資料使用，實現消費者賦權，強化市場紀律

故開放銀行很可能是為了彌補市場競爭不足的狀況下所催生，第二個是希望消費者能有更多選擇，因為有更多的選擇消費者才會有更多的服務與福祉。可能是習望活化資料的使用實現消費者賦權，甚至是市場機率。故每個國家推行開放銀行的動機與目的不一而足，英國是基於市場競爭的考量，澳洲基本上也是基於市場競爭的考量，澳洲有四大銀行，而澳洲當初是由一個類似中小企

業部的組織所發動，想要針對澳洲人使用數位資料的可及性，以及數位資料如何與消費者間產生關聯，做了普遍性的調查。最後做出的是結論，每個澳洲人應對自己相關的數位資料有主控權，進而推動開放銀行。新加坡推動的動機不太一樣，新加坡的動機可能是想成為全球智慧金融中心，故每個推動的策略不一。故在談及開放銀行時，資料開放類型不同，資料的範圍、種類不同，開放銀行推動的目的也不一樣，推動的主體也不一樣。

開放銀行對業者、消費者而言，這是一個數據授權的過程，故必須有授權的兩端，意味著什麼？以前談及下圖時會認為，如果詢問銀行，倘若將左下角的巨型科技業者掩蓋，詢問銀行是否要從事開放銀行，銀行會說不，因為沒有道理資料從多的一方到少的一方，尤其是流向潛在競爭者。故資料的流向會是單向流動，從金融機構流動到金融科技業者時，絕大多數銀行會抗拒。故在法律制度的設計與選擇上，有的國家會採取法律規範強制的做法，如我們談及 PSP2 時，會發現針對支付業者事實上是法律規定必須做的，對於英國而言也是一個法律規定，以澳洲而言推動更極端。惟如新加坡、香港、台灣不是這樣的做法，該做法是認為不應該立法強制，而是由業者決定是否運行，但會產生如業者只看到上半部的圖像會有抗拒，如將下半部的圓提供給金融服務提供者，可能很多金融機構會接受。如果巨型科技，像是 Google、Facebook、Amazon、Microsoft 等，能將他們擁有的資料流動給金融服務提供者，金融服務提供者多數會接受。故今天題目定開放銀行與消費者賦權的想像是，當資料在不同行業間進行跨業流通時，可能對消費者帶來便利、選擇甚至是補給，可以帶來某種想像，消費者自此取得對自身攸關資料的主控權，由消費者以資料可攜的方式讓自己的資料在不同服務提供者間流動，最後結果是讓消費者的選擇變多，這是開放銀行可能可達到的想像。這個想像有個國家正在做，即澳洲。

資料來源：臧正運, 開放銀行的關鍵挑戰－第三方服務提供者之治理模式選擇, 財金資訊季刊, 第 97 期

三、發展模式

（一） 開放 API 框架模式

第一種模式是，政府基本上鼓勵但不使用法規強制，以幾個做法鼓勵。第一，鼓勵業者自主開放。第二，頒布時間表告訴業者什麼時間該做什麼事，第三，篩選出候選應用程式告訴業者，哪些類型的 API 是這個市場上消費者可能會需要的 API，以新加坡為例就做出類似化學元素週期表的 API，列出 411 個不同 API 的應用方案，告訴業者如果是業者想要推動開放銀行 API，就請以這 411 個為主要標的，也很有可能是建置一個網站或是程式介面的註冊庫，或是針對第三方服務提供者如何與銀行往來間做建議與管理。故沒有法律規範的強制，但政府透過很多措施提供框架，鼓勵業者自主開放，這是一種發展模式。

（二） API 管理中心模式

第二種發展模式較像英國，甚至於像歐盟現在所進行的事，成立 API 的管理中心，由政府新設或委由一具公信力的機構，負責協調與制定 API 標準與資料交換格式，並設計一套治理架構，處理消費者與銀行、銀行與第三方服務機構以及消費者與第三方服務機構之間的關係，及頒布 API 上架的相關流程與規範、確保資訊安全的管理機制，未來爭議發生時，有無爭端解決機制。故這是開放 API 管理中心的模式，台灣比較像是 API 管理中心模式及前所提的開放 API 框架模式的綜合體，因為台灣有許多特色的周邊單位，其中包括財經資訊公司扮演類似這樣的角色。

（三） 標準制定機構模式

所為標準制定機構模式是成立一個標準制定機構，由該機構將所有利害關係人集合共同制定開放 API 的標準，不另外成立平台，不另外設立管理中心，基本上是為了定標準而設立。澳洲一開始是這樣，由聯邦技術研究院角色為政府機關，在該組織下有個智庫組織為 Data 61，相當於澳洲版的工業技術研究院。請 Data 61 集合相關人員成立委員會，訂定開放 API 的標準。不過澳洲現在也開始走向 API 管理中心的方式。

四、第三方服務提供(TSP)模式

這是簡單幾個國家目前發展模式，所有環節無論如何發展都落入一個最根本問題，即如何處理銀行端、消費者端與第三方服務者端之間信任的維繫，以及記憶體系可能產生破口。其中最關鍵的是對銀行而言，與其往來、資料託付對象，也就是第三方服務提供者應如何管理、監理。故我初步作四種類型化，這是國際上常見的狀況。第一種是，將原本施加在銀行的作業委外管理規範，繼續套用在銀行與開放 API 的第三方服務提供者上。好處是主管機關監管方便，只需要使銀行負責，由銀行承擔其與第三方服務提供者往來生命週期中的所有風險，以及最終面對消費者的責任，我稱此為訴諸作業委外監理規範，優點是易於管理。缺點是，銀行不會自己承擔成本，故會將成本轉嫁給與其合作的第三方業者，最後造成的結果是銀行只會選擇夠大的第三方業者往來，最後的狀況可能是，市場上需要多元的消費者沒有辦法出現，因為銀行沒有足夠的誘因與第三方業者合作，所有責任被加諸在銀行身上。

有另一種模式稱為委由 API 管理中心把關，優點是第三方業者不用磋商，有一個場域可以測試與驗證，資安與隱私保護的標準可以被標準化、大量推動，但很有可能對特定群體不利，因該 API 管理中心由何人中控制、掌控不清楚，何人可以參與也不清楚。還有一種狀況是制定並適用產業自律的標準，由產業自律，由業者設定自律標準，好處是可以設定最低標準，使業者做彈性調整。缺點是產業自律很容易流於形式，且產業自律的問題是，請問是什麼產業？以台灣為例，台灣現在基本上是產業自律加上委外作業規範綜合監理，所以產業自律是由公會即銀行公會，頒布第三方服務提供者往來相關規範，而此產業魔有包括到第三方服務提供者，第三方服務提供者在過程中沒有代表性，沒有發聲者、話語權，很有可能意見被忽略。最後是由主管機關直接納管，當主管機

關直接納管第三方服務業者時，銀行會較舒適的與第三方服務業者共同互動，因為無需承擔最後責任，會有意願互動，且較可有效落實消費者與金融安全的保護。缺點是，是否有法源依據、足夠的監理資源，第三方是否有足夠的成本擔負未來相關的法遵支出，以上是幾個類型化概念。

資料來源：臧正運，開放銀行的關鍵挑戰－第三方服務提供者之治理模式選擇，財金資訊季刊，第 97 期。

五、我國開放 API

（一） 發展進程

台灣是一個自願且自律的方式推動開放銀行，分三個階段。第一個階段是公開資料查詢，第二個是消費者資訊查詢，涉及產品資訊與帳戶資訊，比方帳戶開戶、附屬業務與信用卡的申請、消費者個資與帳戶的查詢。第三個階段是消費者可以要求第三方服務提供者直接發動支付指令給往來銀行，據此啟動某個支付交易。台灣目前在第二個階段，惟第二個階段尚未正式上路。

（二） 監理架構

1. 中華民國銀行公會會員銀行與第三方服務提供者之自律規範（銀行公會）

台灣推動開放 API 的監理架構是因為要處理銀行與第三方服務提供者之間的關係，故以銀行公會所頒布的自律規範處理此關係。

2. 開放應用程式介面（Open API）技術標準規格文件（財金公司）

開放應用程式介面會涉及技術標準與規格，故需要有人提供，故財金公司提供。

3. 金融機構與第三方服務提供者辦理開放應用程式介面（Open API）業務安全控管作業規範（銀行公會電子化委員會與財金公司）

營運過程中會有業務安全控管的作業規範需被滿足，故此時會是銀行公會電子化委員會與財金公司一起提出安全控管的作業規範。

這是台灣現行推動開放銀行的監理框架，我列出幾個自律規範的重點給大家參考，此資料為政大金融科技研究中心過去所提供，但非金管會最後的版本，僅供參考。該自律規範據說金管會已經核備。從中可以凸顯出現推動開放銀行的困境，第一個問題是會員銀行與第三方服務提供者進行相關業務合作時，除應遵循相關法令外，應依照本自律規範辦理。第三方服務提供者不受金管會監管，銀行可以以契約關係監管，要求銀行先遴選往來的第三方服務者。第一個標準是資本額與營運資金規模要相當，不可經營非法業務。第二個該團隊需有穩健經營能力適足之經驗及專業能力。第三，應具備網路安全與資訊控管之風險管理能力。第四，應符合會員銀行與第三方服務提供者業務合作應適用之開放應用程式介面技術標準規格及業務安全控管作業規範。第五，第三方服務提

供者之負責人及經理人應無「銀行負責人應具備資格條件兼職限制及應遵行事項準則」第三條第一項第一款至第十二款所述情形，並出具相關之聲明書。似乎等同於，使用規範金融機構的標準，規範第三方服務業者，否則為何需要負責人、經理人應出具應遵循事項所規定無消極資料要見的聲明書。這是目前監管的現況，也因此實務上許多合作停滯。

第三方業者應遵循洗錢防制法、資恐防制法、個人資料保護法、消費者保護法及主管機關所訂定之相關法令規定等及本會相關自律規範之要求。應遵循本自律規範、訂定標準作業程序，執行消費者權益保障（包括消費者資料保密）及採取符合第四條第四款所規定技術標準規格及業務安全控管作業規範之相關措施，以確保資訊安全。不得違反法令強制或禁止規定、公共秩序及善良風俗，且不得有侵害會員銀行、消費者利益或其他不當之行為。第三方服務提供者利用消費者資料為行銷時，消費者表示拒絕行銷者，第三方服務提供者應立即停止使用該消費者資料行銷，並至少於首次行銷時，提供消費者免費表示拒絕接受行銷之方式。

重點是要求會員銀行簽訂與第三方服務提供者簽訂業務合作契約，且有規定應訂定事項。等於是透過銀行公會頒布的自律規範，間接將第三方服務業者納管，只是該納管仍將管理責任賦予會員銀行，使銀行承擔最終責任。故要求告訴第三方服務業者應遵循的事項：智慧財產權之歸屬及使用、第三方服務提供者對外行銷時應遵守之事項、消費者爭端解決機制，包括解決程序及補救措施、會員銀行與第三方服務提供者發生爭議時，其爭議處理機制、與第三方服務提供者終止業務合作契約之約定、若發生異常事故、重大缺失或違反法令之情事，第三方服務提供者應立即通知合作之會員銀行及消費者，並應立即採行緊急應變措施以降低對合作之會員銀行及消費者可能造成之影響、第三方服務提供者如有違反上述各款任一規定，致損害消費者或合作之會員銀行權益時，應負損害賠償責任，合作之會員銀行並得終止合作契約、消費者資料保護與爭議解決。故於消費者與第三方服務提供者或會員銀行發生消費爭議時，會員銀行應提供消費者申訴管道，並應提供消費者協助，以妥適處理消費爭議。會員銀行應與第三方服務提供者約定，如消費者向會員銀行提出因其與第三方服務提供者所生之消費爭議而受有損害者，除會員銀行得證明消費者有故意或過失者外，於一定金額內由會員銀行先償付消費者，再依業務合作契約之約定向第三方服務提供者求償。簡單而言，藉由銀行承擔第一線消費者的賠償責任，對銀行而言較困難，對第三方業者而言合作也會有困難。

伍、消費者賦權的想像－以澳洲的消費者資料權為例

一、澳洲消費者資料權的定義與核心

如開放銀行可以做到前述的這些事，我用一詞總結稱之為「消費者賦權」。即消費者可以透過科技的某些方案與法律基礎的設施，在消費者的主控與同意之下，選擇，與其有關的資料應在哪些服務主體之間流動、被使用。澳洲將這件事情推動的十分極致，他們認為不只要推動消費者賦權，還要在法律上賦予

消費者資料權，稱為 Consumer data right（CDR）。有部法律為競爭與消費者法，在許多英美法系國家競爭法與消費者法的主管機關為同一機關，法律權源也是一致的，故可透過修改一部法律，將某種權利、義務進入法律規範中。何為消費者數據權，是一橫跨所有經濟的改革，漸進性採取不同產業推展。目標是消費者資料者主體可以效率且便捷接近與自己有關的資料，而這些資料是長期以來被公司所控管的資料。並且可以使消費者要求數據持有者將其所管控的資料以安全的方式揭露給第三方，或是消費者本人。

二、澳洲消費者資料權制度的監理分工

故在澳洲的脈絡下，只給予第三方業者資料讀取權利，而非啟動交易的權利。此權利為一新的權利，故需要法治工程建構。其中最重要的法治工程是，請問誰可以當合格的資料接受者（accredited data recipients）？這些認證條件包含隱私、資訊安全的要求，由 Australian Competition and Consumer Commission（ACCC），是消費者保護與競爭法的主管機關，在澳洲制度下推動開放銀行的不是金融監理機關，而是競爭與消費者保護的主管機關。透過機關協作，假設有個被認證合格的資料接收方違反義務，ACCCC 會給予行政處分暫停、撤銷、停止服務。但中間會需要監理分工，故有三個重要的監理主管機關必須參與，一個是 ACCC，依照競爭與消費者法所立的消費者資料權頒佈授權的法規命令，稱為消費者資料權規則，該規則已頒布，負責擔任認證資料接受者角色，並建立與維護註冊合格資料接收者。故可以在其網站中查詢哪些是合格的資料接收者，並確保各方遵守相關法規。但在此過程中會涉及個資隱私問題，故請澳洲隱私保護辦公室（OAIC）參與，其負責制定消費者資料權的隱私保護準則，且監督遵守關於隱私投訴，這是他們的分工。至於資料標準機構（DSB），就是 Data61 所設立的資料標準機構，負責制定開放 API 當中的技術標準。

三、澳洲消費者資料權制度－資料接受者的認證條件

第一需滿足適格要求，何謂合格的人，可能包括資本額、團隊經營管理。第二，需重視資安的相關措施。第三，須重視爭端解決機制，且該爭端解決機制應是內、外部都有，所以銀行端要有爭端解決機制，第三方業者也需要有爭端解決機制，且此兩種類型的業者都要有客觀適用的外部爭端解決機制。最後，需要有保險，因澳洲很特別，澳洲有資安險，英國的也有資安險。倘若第三方業者是新創公司，要求他們作合格的資料接受者，承擔所有資安外洩、個資外洩的責任不切實際，需要保險制度的搭配，故需要這樣的商品與服務。

四、澳洲效費者資料權制度－同意權行使

除此外，消費者賦權關鍵在於消費者有主導與主控權，如果消費者沒有主導與主控權所有都是虛無。故他們十分重視同意權行使的規範，而列出幾個同意權行使的準則。第一同意要自願，明示、知情後的同意，每一次同意都是基於某種特定目的。授權要有時間限制，每次授權不可以超過十二個月，台灣銀行公會的版本是每次授權是不超過三個月，消費者要能夠很輕易地將其授權撤

回，消費者應有權利選擇其 CDR 被刪除，此概念類似被遺忘權。同意權管理的機制，規範上要能夠實現同意權的管理，制度、介面設計上也要能夠實現同意權的管理，故要求銀行與資料接受者製作 Consumer dashboard，消費者可以拿手機準備授權時透過該儀表板，知道過去一段時間做了哪些授權、基於何種特定目的的用途、妹每次授權期間多長、何時曾經撤回、授權的對象為何，透過儀表板一目了然，這是他們所說同意權管理的機制。澳洲的要求是銀行要有此儀表板，第三方業者也必須要有。這是我認為台灣在討論開放銀行中缺少的一環，我認為很重要。

陸、消費者賦權的挑戰

消費者賦權會遇到何種挑戰，以下列舉幾個問題。未來會有合格資料接受者，且該合格資料接受者，假設資料進行跨產業的流動，會有不同產業的合格資料接收者，應如何設計合格資料接受者的監理機制。光是在金融領域就不容易回答，以開方放銀行為例，現如何納管第三方服務業者，現沒有法律依據。實務上問題是，即便要納管第三方服務業者，主管機關的擇定由誰擇定？第二個遇到的挑戰，如何推展消費者資料全的法制化工程？所有消費者賦權的想像都建立在重要得前提，即消費者是基於充分自願知情同意的狀況下，選擇與其有關資料在不同的服務主體之間流動。在法制上賦予消費者這樣權利，需要一個法治工程，該法治工程如何落實。以台灣為例，如今天賦予消費者資料權，需要修改何法？銀行法？金管會組織法？金融消費者保護法？消費者保護法？公平交易法？如何入法？何法？哪個主管機關？這是挑戰。

有了法治權利，會需要實質上的管理機制，特別是同意權的管理機制，如同討論個資隱私保護時，會討論資料可攜權。資料可攜權是一個理想的境界，惟該資料需可攜，又要確保每次的授權都是消費者同意、每次授權都沒有逸脫特定目的，有無可能有制度面或是設計面上，可以幫助我們達到的。故有無法律規範、技術方案的問題，而技術方案與我所說的助推設計有關係，即有沒可能如同澳洲有 dashboard，可以清楚幫助消費者做消費者同意權的管理。如果知情同意可以透過技術做到這種程度，如果我們在制度或介面上使用助推機制，是否可以更便捷幫助消費者行使同意權。在其中有無助推設計的考量、技術方案的選擇可以是什麼？所有與技術方案的運用都會牽涉到的問題是，技術會出錯，需要有外部的集合，誰能做外部集合、如何運作？更重要的事，可以很理想的說要求大家做到完美的技術方案，惟成本誰承擔，銀行？業者？新創業者？還是轉嫁給消費者？

最後，同意不等於主控，資料賦權不等於資訊安全。有再好的同意權管理不代表資料被我處理跟利用擁有絕對的主控權。資訊安全一旦出問題所以的基石會被推翻，應由外部監理？何機關監理？還是市場有無紀律發生的可能性，所謂市場紀律就是消費者本身，台灣在金融領域，談太多金融監理上著重金融機構健全業務經營、金融體系的穩定、金融消費者的保護，惟金融消費者如何在過程中自負其責，這些是我看到的問題思考與挑戰。

問題與討論（敬稱省略）

李建良：

所謂第三方服務業者為何？若從法律觀點而言會是什麼。第二是消費者資料權利，如果將消費者去除，只看資料權利，但我們通常是說個人隱私或個人資料保護。這個概念中想要承載的權利，當然最後談及所謂合格資料的接收者，已經將此權利的形貌畫出，追根究底而言，該資料是自己的還是別人，資料的受詞是誰，如從接收者的角度而言似乎是他人的資料，是否為新的資料權的形貌。

臧正運：

第一個是，現在第三方服務業者及合格資料接收者，現在我們的想像會像是個法人。以台灣為例，譬如麻布記帳、CEmoney，可能就是第三方服務業者，原本可能是科技業者、製作應用程式。也有可能是較大的公司，如電信業者、證券集保結算所。故第三方也可以是個法人，對他而言只要能夠透過取得客戶授權下銀行端或是他業而來的資料，針對資料加值提供某些服務或多元的選擇都可以當第三方服務提供者。才會有討論是這種類型的人看起來也承擔大家的信任，拿資料的人被我們託付也是重要的東西，雖然在想像中談及錢與資料時會認為錢比較重要，承擔資料的保管者、處理者其實是很嚴肅的，只是現在沒有明確的機制一定要有何條件才能做這件事。從開放銀行的領域目前有的討論只會至少要確定一定資本額的要求，至少要有專業度，該專業是團隊有處理資料的人。目前看到會是透過資格要件的符合，只要符合這些資格要件就可以作為第三方服務業者，前提是銀行需跟你有往來。在法律上我會較傾向認為第三方服務業者是一個法人，個人承擔有無辦法承擔此角色，我認為相對而言是較困難的。

第二個問題，權利的本質為何，我認為是個大哉問。我隱約看到他們在討論這個問題的脈絡是，他認為隨著科技的發展公民有越來越多的數位資料被產生，該數位資料其實會對某些業者帶來利益，惟該利益不一定會反饋到公民本身，這件事法律上應如何評價或看待，因此澳洲慢慢發展出你應該有個權利，惟是否應該這麼快拉到權利層級，我仍在觀察。他們打算先做銀行，銀行做完做能源、最後電信，三大特許行業做完後心中的想像是可以跨業的資料流動，因為才剛開始，我也不確定會如何發展。

李建良：

第三方有可能是資訊業者，聯想到 Uber 問題，Uber 不認為他是運輸業者是資

訊業者，交通部認定其為運輸業者而處以罰鍰。這種情形會不會相同發生，即主管機關認定是金融業者，而非資訊業者，故以金融業者的監管方式直接進行管理？

臧正運：

我認為有可能，現在這種間接式的管理有點這樣的影子存在，變成是要求銀行確保這一方的業者更像銀行，確實會有這樣的問題。

邱文聰：

今天報告的題目包括開放銀行、消費者賦權，聽起來是在談兩件事，如果從打破銀行壟斷讓金融產業更有競爭性的角度討論此問題，我們看到現在台灣做法透過銀行公會的自律規範，由各個會員銀行與第三方業者合約訂定，似乎不能達成該目的，因自己找他人合作要打破自己壟斷或寡占的局面，似乎不合理。台灣自己在發展開放銀行背後的思維為何，與英國、澳洲似乎不一樣。

第二，與消費者賦權有關的是，我認為很有趣，原因是該問題不只出現在金融消費市場上，同時會出現在其他場域。我目前在做的計畫是與醫院，醫院也擁有大量的病患資料，這邊出現相同的問題是，每家醫院都將自己所握有的病患的病歷資料當作是投資很多努力下的記憶，為何要與別人分享？會有一個這樣的困境。我們的計畫也意識到這樣的問題，現在有越來越多的資料使用需求，希望將 A 醫院的病患資料與 B 醫院的病患資料整合做加值應用，就會遇到其實病患自己的資料控制能力有限，一種模式的提出會與剛剛臧老師所提到的消費者的賦權模式很像，即找一個數據經紀人或是第三方幫病患主張他的權利，故病患可以將其隱私偏好與第三方分享，他願意在何種情況下，為何種目的提供其醫療資訊做何種應用。這個第三方會有很多不同病患給予這樣的授權，資料使用方可能是未來其他醫院或是其他想要使用病歷資料進行醫學產品開發的數據使用者，就可以與第三方業者要求病患資料。由第三方業者授權，某 A 病患在 X、Y、Z 的資料是可以被釋出或是不能被釋出。這套想法其實與臧老師提到的消費者資料權是相似的。這樣的風潮其實已經在許多不同領域出現，台灣目前有嘗試在做同意權管理形式，透過資訊系統研發，類似剛剛所提到的消費者儀表板的模式，他們比較想做的是動態同意模型，可以隨時動態的控制資料利用，故當然需要一個儀表板可以有效管理醫療資訊。有關消費者賦權這塊在跨領域已經有些實踐可以參考，但是前端有關開放銀行的部分，想要達成的目的為何，我較不清楚。

臧正運：

第一個問題，台灣有無推定開放銀行的政策思維，我認為應打上一個問號。可以很清楚的看到其他國家推動的脈絡可循，台灣的脈絡有點像是因為別國家推定，會有些討論。其中一個討論是台灣的零售銀行市場與其他國家的結構一樣嗎？如英國與澳洲都是幾大銀行獨占的市場，台灣不是。台灣是某些實證研究會認為像是獨佔性的競爭市場，也不是完全競爭，又有人認為是過多銀行。故台灣業者一開始很抗拒這個想法，會認為已經夠競爭了，為何還要將競爭帶入。從另一面觀之，我認為主管機關帶進競爭非其本意，惟其同時也認識到數位金融與金融的浪潮可以讓消費者有多元選擇，故他現在在尋找一平衡點，如何不要過度改動既有規定，可以使消費者選擇看起來變多。故台灣的背景是一連串思維下的結果，很難法規強制，當然與我們的結構限制有關，即要使用法規強制應改什麼法，有何既有的法規命令可以授權，可能找不到。如找不到要修法，應修何法？不知。故在政策上合理的選擇是由業者自願自律來做，惟自願自律會碰到的問題，會造成如業者不願意應如何？故台灣最近 My data，是政府各個機關管控的資料，由我作為一個公民，我使用 My data 的資料，如財稅資料、所得資料、勞保申報資料，與 My data 有往來的十間銀行，申請信用卡、房貸、車貸。這與開放銀行要做的事有點像，都是給予消費者資料可攜的選擇，會發現這些銀行有的在做開放銀行，有的不做開放銀行，先進入 My data，因為 My data 的推動者是國發會，金管會比較慢。金管會現在分階段推開放 API，但是具體規範要到年底第二階段才會正式上路。在正式上路前，他選擇十八種不同服務內容的 API，跟大家說第二階段可以做十八種 API 的內容，在正式大規模開展前可以做，但是因先至主管機關做業務試辦。會導致金融業可能想做，但又發現有規定在，做起來不會如此快。第三方業者想要加入又發現金融業的門檻如此高，以資安而言，必須要有 Iso27001，但許多新創業者沒有，故需花成本。新創業者無法進入，金融挑大的公司合作，但大的公司又會認為為何要給金融業者主導，台灣會變成較詭異的呈現。

有關第二個問題，其他領域的師長也有在做類似的計畫，如剛剛所提的第三方業者是一個資訊中台或是隱私權管理工具，也許背後可以搭建區塊鏈技術，有些醫療機構有在做類似的事情。我個人認為這件事會是趨勢，惟此趨勢能否真的解決，確實做好同意權管理，很難管理同意權管理完後發生的問題。而這段應如何解決，目前我沒有答案，感覺技術方案尚未到達。

邱文聰：

有關 My data 的部分，銀行的動機為何？基本上如果是銀行就是公會的會員，也因此他可以取得聯徵中心的資料，為何需要靠 My data 才能核發信用貸款？

劉靜怡：

聯徵中心的資料有那麼多嗎？如剛剛提到勞動部。資料的多樣化可以讓洋行的決定更佳精準。

林建中：

最可怕的是財政部的財稅中心，你的利息、收入、股利、股票、所有繳的稅、房地產等。所有資料中最可怕的是國家將手伸到每個空間中。該機構有兩個進入的管道，第一是拿到勝訴判決，可以要求看債務人有何資產。最常出現的問題，當有人死亡，繼承人可能不知道被繼承人有多少遺產，希望可以瀏覽。其他金融的三個領域，第一個聯徵中心查的是信用問題，有關財經，基本上是銀行維繫、內部的部分，還有集保中心主要是管理股票，有許多證券投資的數據在那。

劉靜怡：

純量的數據，即 A、B、C 股票各有多少。我好奇的是這種資料會存多久？比如將某一檔股票賣掉，資料會一直留存嗎？二十年後還是可以查到某年某月某日賣出何股票嗎？

林建中：

基本上沒聽說過有刪除。

臧正運：

除了股票、有價證券的資料，還有基金資料、所有有價證券的資料。

林建中：

故能想像到的基本上所有都在政府手上，雖然有些是用財團法人的型態，有個是用財政部的型態，有個是用公司的型態。現在可以至聯徵中心索取個人信用報告，在貸款時銀行常常會問所得稅申報資料、扣繳名單，是聯徵中心沒有的。而最後其實是會到銀行，政府已經將手伸入。這是我第一個問題要做開放銀行期待的利益是什麼？銀行就算沒有辦法自行取得，也可以要求客戶提供，且最後客戶也會提供。推動開放銀行主要的利益為何？只是多一組人擁有我們的資料嗎？於我而言，知道哪張信用卡較好用不是一個很好的誘因。

臧正運：

聯徵中心擁有的資料有極限，My data 的資料如財稅資料、勞保加保的資料，如同劉老師所說的，他可以透過這些資訊針對你做更精準的分析，決定要給予何種條件，這可能是其中之一。另一個會是較商業的考量，My data 有參與，開放銀行有參與，對於擁抱金融創新的形象是好事，我認將像是這種考量綜合的結果。台灣推動開放銀行的利益為何，我個人認為是，如同老師所說的使用信用卡，我認為這不是最有力量的場景，最有力量的場景對某些人而言也許是單一財富管理的介面，對某些人而言是清楚知道這些事是有幫助的，如如何消費，將錢花費至何處。

林建中：

使用者是知道，只是為了克服這種麻煩創造一組人再將這些資料擁有，其實保險法領域也要討論要有保險資料中心。基本上已經有銀行、所有的投資證券、保險、財稅中心在掌控所有的想像課稅的東西。我懷疑的是利益有這麼大，Google 勉強還不知道的可能是所的存款，也許知道。如趨勢是對於資料更保護，開放所創造的可攜性的好處我不認為有足夠大。因為可攜性是現在即可做到的，其實我們並不是沒有選擇。

臧正運：

開放銀行的場域是一件事，我個人認為像是跨業的資料流動可能是另一件許多人在討論的事，確實可能會對生活帶來許多改變。

楊岳平：

第一，很多人並不是很在乎隱私、個資，既然已經有很多人擁有我的個資，再多一個人也無仿，對於很多人而言個資比率不見得有如此高。我個人認為最近大家在討論的一個場域，即跨機構的支付或是交易，這是很多人在討論的事。像是現在在通動的電子支付，希望辦了一個 Line pay 後，Line pay 可以付街口支付等，都需要某一種串連。開放銀行在想像的是讓所有交易由金融機構串連，讓所有交易是一站式到位，利用此窗口做所有的交易。這對大家而言，好處較多還是壞處叫多，我認為是取決的每一消費者對自己個資的重視程度，以及對於金融便利性的重視程度，背景大概是如此。反之，我不認為台灣的主管機關推動開放銀行是為此，另外有個政治因素在背後，是新創業者推波助瀾的結果，近幾年新創業者的政治聲音很大，使得金管會必須做回應，許多金融科技方面的回應是為了做到某種對於新創業者可交代的程度，是半推半就的結果，金管會會打從心底不想做，會造成行政負擔，才會有如此多奇怪的模式。

臧正運：

其實關鍵是，如果剛剛岳平說的多一個選擇，是否要使用可以自己決定。這也是澳洲在推開放銀行原因之一，澳洲在推動開放銀行時做裡一個國家以外的調查，該調查找了一個律師主導，他們最在意的是有無選擇，確實這些國家是否有做非常嚴謹的成本效益的分析，我認為似乎沒看到有任何一個國家是這樣做的。

林建中：

我懷疑其他國家是否有如同我國如此密集由政府資助的資料集中化，在台灣這件事已經做了很久，這件事其實他們已經在我們不知道的時候就在做，是否會有背景上的落差，不太確定其他國家是否有如此徹底的規劃資料庫、跨業者、跨級別的規劃。

蘇凱平：

我認為將資料集中在一起，我能想到這件事的能獲大最大利益的人是檢察官，可能是從旁論的角度。最近科技偵查法的草案停滯，如果要提出其中可能會有較大的改變，以科技偵查法的觀點觀之，是十分需要一個將資料集中、跨場域的處所，如能將全部合一，而非散落在不同的機構，這件事對於檢察官而言是很麻煩的。

黃相博：

檢察官有自己的內部的體系，他們已經有可以串連的平台，不需要靠外部開放，可以連線所有財政部的資料。

蘇凱平：

我在想應是金流，比如洗錢犯罪，洗錢犯罪確實會有找到各種金流的管道，但這件事他們必須自己做勾稽。且洗錢犯罪在法院審判遇到最大的問題是，就是起訴時講述金流的流向，而在審判中辯護人講另外一個故事，判決認為不是這樣勾稽方法不是這樣。如果今天是以開放銀行或是類似的方式，可以將資料用很整齊的方式勾稽，對於我而言，如果我作為審判者，是一個特別有說服力的方式，編排方式應該是要如此的，因為國家已經整理好給我了，這是第一個。另外剛剛提到的，API 厲害的是，我每週都用一兩次使用手機付台北市路邊停

車費用，故會勾稽的已非僅金流。

劉靜怡：

Line 提供的所有生活付費功能，基本上就已將生活的軌跡全部勾勒。

蘇凱平：

以前要跟蹤需要用 GPS 的方式，惟因通訊保障監察法有非常多設計的機制反抗，要得到 GPS 的偵查不容易。如果透過 API 的方式，連結到很多物理性，如付水電費的金流勾稽是一回事，惟如同靜怡老師提到的，在某處的 711 用 Wallet 付款、在哪停車等，我的行動軌跡是很清楚地被勾稽，這件事對於國家而言有巨大的利益。

林建中：

我們到底有多少人是罪犯？

劉靜怡：

對於治理者，無論是公部門、私部門，有一個基本心態是，可以掌握的資料越多越安心，因為需要時即可使用。最近有一群台灣的資料科學家，大部分是以經濟學家為主，加上一些社會學家。他們與政府部門談好，號稱要用行政資料放置在一平台上，想要做研究時，就透過該平台拿到資料，事實上是這樣運作。甚至包括最近談及從小學生開始到高中要製作學習檔案，這個東西就是裡面的一部份。可以想像學界都可以這樣做，其他部門的誘因其實更大。

楊岳平：

剛剛文聰老師提及，於醫療領域中有數據經紀人，我的理解是有點像是一個代理人，像是不動產仲介只是變成是醫療仲介，做此有需要特別許可嗎。

邱文聰：

現在沒有管制。

楊岳平：

也不會被當作是醫療產業？

邱文聰：

不會，就是做醫療資訊，即個人的管控。

楊岳平：

但其時手上會有很多個資。

邱文聰：

他的模式是他只透過偏好

楊岳平：

偏好也是個資。

邱文聰：

偏好只是，如我的血液、血脂的相關資訊，同意做高血壓的研究。事實上血脂多高市不清楚的。

楊岳平：

如剛剛李老師的擔憂，這樣資訊業者會不會如同 Uber 般被歸類為某種特許行業。

邱文聰：

會與剛剛所提及的金融第三方服務業者不太一樣的地方是，第三方服務業者需要知道資訊內容，以便幫忙做管理或建議，故程度上有差別。也有些醫療的第三方業者是真的拿資料的，但現在的模式是可以完全不碰資料，只有如透過區塊量的方式記載偏好，所有的要求進入，他幫忙過濾後，是作為一個開關的角色。

楊岳平：

但他拿取你的資料又如何，也是基於同意下給予的。

邱文聰：

他不想負擔這個責任，故他認為不要給予資料。

劉靜怡：

對他而言，目前要發揮的功能這樣就夠。

邱文聰：

且區塊鏈基本上會有資料量的上限，不可能在每個區塊上寫太多東西。寫偏好就夠了，不需要將完整病例資料放上。

高國祐：

剛剛提到當資料安全發生問題，造成消費者損害時，是由銀行負第一線的責任，代表澳洲規範設計預想由銀行承擔風險，為何在這種情況下不是由資訊持有者，即平台業者負責。舉例而言，在使用麻布記帳時，條款中有寫到如因為資訊洩漏導致損害，明白揭示不負損害賠償責任，但是與銀行的條款中也明示如果將自己的資料給予第三方業者造成自己的損害，銀行也不負責。規範設計上風險應如何分配？

洪于庭：

我認為在全球應該沒有任何一企業比 Google 擁有我們更多的資訊，事實上在 2013 年美國有個判決揭示當信件進入 Gmail，Google 會掃瞄信件內容，甚至是刪除的信件，Google 敗訴原因是因為未於隱私權條款中揭露。銀行與 Google 時，如同國祐所提及的資料洩漏問題，Google 會因此承擔何種責任嗎？或是可以獲得賠償嗎？畢竟是 Google 使用者的資料被洩漏。另外，對於 Google 而言，有何誘因跟銀行合作，沒有一家銀行的規模是可以與 Google 所匹敵的。

臧正運：

剛剛所說的脈絡，分配責任的機制不是澳洲，剛剛所說的是台灣的現狀。看起來台灣的自律規範會是由銀行承擔第一線責任。為何會是這樣，也就回答你的問題，銀行與客戶簽約，銀行表明不負責任。與第三方簽約，第三方也表

明不負責任。在台灣銀行是無法逃脫責任的，因為金管會。就有人舉例，如果今天在某家銀行提領十萬元，在門口被搶劫，銀行可能都還要賠償。這是一個我們一直以來對金融消費者保護的思維，才會變成我們都認為銀行最安全，最後制度懲會如此。澳洲是規範面上說得很隱晦，依照事件可歸責的對象咎責，但沒有揭示銀行須先賠償，這是澳洲現在的處理模式，即個案判斷。但是如何操作因沒有看到實際上發生問題。確實如我所述，責任分配是個難題，光是判斷責任分配，金流支付，哪個環節資料被，技術方案，資訊流的環節，光是要判斷責任分配要先找出問題歸屬，今天如是 Read Access 不涉及金流指令的啟動，關鍵會是在哪個環節資料被外洩，如何證明。故即為何我會說同意權管理的機制是否有技術方案可以記錄資訊流的環節，如果有可以記錄資訊流環節，才有辦法做理想的責任分配，否則，最符合政治正確的考量會是使最有能力承擔的人，負最後承擔的責任。至於 Google 問題是十分難解，我猜想如果有誘因會是階段性誘因，可能在某階段會想在某個市場做某件事，會需要金融服務的執照，對他而言沒必要，可能會先選擇合作，這種階段性、策略性的做法。

黃相博：

在開放銀行的環境下，是有意將消費者賦權的概念下只考量消費者，是符合金融消費者保護法或是消費者保護法的消費者概念嗎？如果不是消費者，在開放銀行環境下有無您提到的資訊權的部分需要被考量？

臧正運：

現在所說的消費者，因為在台灣的情境是必須是由會員銀行給提供客戶資料給第三方服務提估中。故自然會落入原本對金融消費者保障的範圍，我心中想像的消費者不必然要受此拘束，可以是廣義的依據在何種連結或情況做資料的移轉跟流動，我自己的想像是沒有特別侷限在一定是金保法上的消費者。台灣能做的部分以自律規範而言，他們的自律規範也是用消費者的用語，定義是接受會員銀行或第三方服務提供者依業務合作契約所提供的金融商品或服務者。

黃相博：

我們講到消費者時有時會有框架，如是銀行客戶有時是公司，在發放薪資，或是貨物交易等，那些帳戶資訊也是有可能會落入。如是在開放銀行的環境下，也許會有資訊交換的可能性，我只是想要確認，是否有刻意要將這兩塊區分。

臧正運：

沒有，我認為可以包含。以澳洲為例，澳洲消費者資訊權的主體也可以是小公司。

黃相博：

台灣的架構下有可能？

臧正運：

台灣應該上會想到那程度，但我個人會認為可以考慮。

楊岳平：

我認為從文字觀察是如此，這裡的消費者定義顯然不是消保法或金消保法的定義，沒有區分並將複雜的排除。基本上只要與銀行有往來就是，可以包括專業投資人。我的問題是不管事消費者資料權利，我們會發現這裡所說的「Data」不一定是個資，其實有很多是非個人資料，理論上是超脫個資法的狀況，至少澳洲所說的消費者的情況是超脫個資法的情況。回到台灣的框架下，台灣暫時沒有立法的情況，唯一會卡住的這些個資法，法人我不確定。法人資料在台灣的情況，似乎是監管真空的狀況，反而澳洲提出消費者資訊權後，好像稍微擴大對法人的保護，至少是小型企業。個資法與資料法是不一樣的概念，資料法更廣，開放銀行或是未來各種開放產業可能會談及的是更廣的資料法，現在連個資主管機關都找不到，更難想像萬一需要資料主管機關應如何。假設有數位資料法或是個資主管機關，基本上非目的事業主管機關，較像是作用法的主管機關，而剛剛正運所提及最後需處理的問題是，第三方服務業者納管，想得更多的是目的事業主管機關，不只是作用法的管理，如果要變成通案，不確定正當性如何處理。

臧正運：

這我也還在思考，剛剛提到澳洲維和會將小型企業納管，我沒有特別看到法律文本，我猜是一種入境依存後的結果，因為最早開始在針對產業做調查的機關是主管中小企業，我猜是因為這意的原因

李建良：

開放銀行的推動背後的思維可能是自發性還是依照法規，代表兩種方法。可能

是業者自己推動，而自己推動不需要理由，這種模式政府或國家需要介入，在進行時會有何種問題,有無可能會有規範真空的問題。而這部分會有問題的是，是否會形成反競爭問題，變成幾家銀行加上第三方業者，形成聯合行為，當國家看到可能須介入調整。另外是一法規的模式，是由上而下的情形，變成國家主動推動，歐盟是這樣，歐盟強勢推動要求必須開放，這種情形必須要有正當性，正當性不足會有問題。

第二部分是否需要納管，這部分與過去的基本想法是組織法思維，好像將依組織成立後就可將所有問題解決。事實上是作用法問題，如環境資源部，在行政院組織法已經有規定，直到現在無法成立，因環境資源部一旦成立，會將所有業務納入，因為所有事都與環境資源相關。事實上作用法的部分是不清楚的，就算成立作用法很難作用。最後，銀行是否有包括郵局，事實上在台灣很難排除郵局。郵局事實

第五次會議紀錄	
時間	109 年 11 月 26 日（星期四）
主題	人工智慧、大數據與個人資料保護
講者	張志偉（銘傳大學公共事務學系助理教授）

內容摘要

壹、問題提出與論述步驟

人工智慧的脈絡此議題在國內談及多，各樣的商業活動及運用已經幾乎想像不到哪個領域是不需要資料、不需要人工智慧技術。包括每日必需的牙刷或是車載的電子系統透過物聯網、網路連結蒐集個人的敏感性資料，或是一般個資，故可以做得部分是十分廣闊。因此，德國聯邦憲法法院於 1983 年提到人被解譯的危險性，以現今而言是十分有可能，如娛樂或電影已經逐步展現，這是個資保護上的疑慮。有些人認為個資保護與自己十分遙遠，不像如同一般的民事、刑事訴訟的侵權有如此深刻的感受，我認為個資保護是將人格權的保護往前挪，個資尚未被侵害的情況下，事先課予責任人。「責任人」的用語，我個人與國內一般用語不太一樣，國內一般是使用控管者、管控者、控制人的角度，這是從英文翻譯而來。惟於德文原文中「Verantwortlich」，是從責任角度，如各位對於資訊行政法在德國的發展，其實是與警察法有很密切的相關。故在警察法上的概念是被理解為責任人，這是我於翻譯上的想法。

談及人工智慧時，不免會提到「ABC」。A 即演算法（algorithm），演算法的部分法律人很難理解，惟於資訊科學而言是很關鍵的，在資訊科學上會傾向於使用學習系統、認知計算取代大數據的用法。在演算法中，我認為較重要的是如何得知，知道如何得知的情況下才能行使個資權利。B 即大數據（Big data），大數據即為今天關注的重點。C 即雲端運算（cloud computing），在雲端運算部分，於個資法中 GDPR 沒有特別提及，基本上我們會將雲端提供者界

定為資料的受託處理人，只要是單純儲存。這三個我認為是在談及人工智慧時不可或缺的部分，一般而言，人工智慧是指使電腦展現出像人類行為的科技，是透過感知學習、推理、協助決策幫助人類解決問題。他的學習來源就是資料，這也就是我們在提及第三波人工智慧會大躍進的原因，因為我們搜集處理資料的能力已經大幅提升。無論是質與量上，都不是以往能想像得到的。故於此集中重點主要還是法律面向的討論，核心的關懷會放在大數據與個人資料保護的面相討論。此面向上會發現悖論的感覺，談人工智慧或機器學習必須要有大量資訊作為基礎，如收集資料越來越多，對於人的圖像掌握、解譯的可能性會升高。表示人工智慧會隨著資料量更深入學習，對個人資料的資料保護更大挑戰，我認這部分是具有衝突的部分。可以看到大部分在談人工智慧影片或是書籍，只要不是法律學門來看，多半會歌頌需要資料收集，許多報紙上投書也是如此。在健保資料的案例上也是如此，我們在強調健保資料庫的運用應更彈性、更開放，背後個資疑慮於企業界中會認為只是杞人憂天。在此面向上主要討論重點要如何探討既有法規範，是否可以或是如何有適應能力問題，要如何解決大數據所提出的問題，在此以歐盟的 GDPR 做為對照的對象，原因除因為是最新個資立法外，另一部份是表現出歐洲對於個資基本的立場，故以此故部分作為探討。

貳、大數據與個資保護

一、大數據運用特徵

可以看到於大數據運用上指涉與人工智慧有很大重疊地放在於，使用巨量資料，透過巨量資料分析技術彙整出，透過分析發現及預測揭露人類行為模式或是各種趨勢，以便於生成新的知識，回饋各個產業的理解。這樣的分析脈絡，與過去不同的是，於一般統計中，原則上希望找到因果關係，惟於大數據運用上，反而強調如何找出不同特徵的關聯性。我們要如何一步一步的去理解，首先要先分析的是資料是否為個人資料，如為個人資料才會落入個資法的討論中。當中也會有非個人資料的部分，非個人資料原則上是指涉事實資歷、匿名資料、法人資料。原則上共通的定義，非個人資料就是無法或是無從識別個人，這也就是匿名與個資很大的區別所在，我發現國內對於「匿名化」，大家的理解好像不太一樣。有些學者理解為假名化，只要有代碼就是匿名化，此部分我認為德國法對於 GDPR 的了解是有落差的。以大數據的特徵而言，被美稱為智慧資料，主要是因為有四個結構性特徵：巨量、資料來源多樣性、分析速度、對於商業、公部分十分高的價值影響。此四個結構性特徵，界定出大數據的特徵。大數據運用的目的有許多態樣，一開始可以界定幾種可能性。第一個是預測群體的行為模式，如針對特定契約的徵信評估、徵信評分的方式，或是從無關聯的資料中累積成個人的數位人格剖析，特定要素長期累積的觀察，如健康狀態、愛好、信用程度，即個人化的大數據運用。數據挖掘或是資料倉儲之用，乃至於科技的偵查措施、追蹤方式，這也是大數據常見的運用方式。

德國學者 Schulz 的分類方式，是將大數據分為以下這三大類，分類方式可

以凸顯並不是每一樣大數據都以個人資料為主。

1. 非個人的模式建構：目的是預測模式、模擬或智能聯網，此類大數據利用之目的並不在於識別出特定個人，因此重點在於如何將投入利用之前將資料匿名化，以及必須投入避免再識別化保障的措施。
2. 用以識別及揀選可得特定之自然人：尤其用以犯罪行為的揭露目的、詐騙預防、保險或財務會計的異常性揭露。就此正當性顯有疑慮：區分公、私部門有別。公部門需另立特別法；私部門詐騙的預防或會計上異常性的揭露，私部門必須投入技術與組織上的要件。在歐盟對於預測犯罪預防性警務人工智慧的使用是有疑慮的，會造成對特定族群或是移民有惡化、標籤化的情形。公部門與私部門在此脈絡下有些不同，如是公部門，會特別要求另立特別法，這是做科技偵查立法的緣由。另一方面，對於私部門，在基本規則前言中，有提到為了要防止詐騙的正當利益，故私部門投入個別、個人的大數據運用，是認為有其正當性。不過做篩選時，是要透過組織上、技術上的框架與條件，還有合法性要求，並且希望做第三方假名化作為基礎，在這些要件下，會肯認私部門也可以做個別化的揀選或是識別。
3. 藉由既有資料的累積、聚集與評價，或連結其他資料，重新生成關於已屬個別化之自然人新的特殊資訊。如個別化的保險模式、遠距醫療或線上購物的助理、車輛追蹤、定位等。著重在後在使用個資時的合法性要件，如同各位熟知的，在第六條中要求首要即需當事人同意，沒有當事人同意的情況下就必須符合第二款至第六款其他的合法事由。需特別提醒的是，在此面向上使用階段上就會有合法性疑慮，故會強制需有有效的假名化措施，並且有透明化的要求，即當事人權利，特別是當事人的異議權，透過此保障當事人。

二、個資保護基本原則與大數據

在此部分有的出發點會與有些學者不同的是，我認為個資的保護是強調資訊自決權而非資訊的隱私權。我對於司法院釋字第 603 號解釋的理解，與國內學界有些出入。該號釋字明白強調的是資訊隱私權，如觀至行文，定義基本上是 1983 年人口普查案的定義，強調其實是自我決定。據此，這兩者的差別在於，強調隱私不可迴避的是必須有合理隱私期待，惟無論在歐盟或是我國的個資法中，資料的定義並未要求要有任何的隱私期待。所有可以識別出、可得特定的個人資料都是個資，未有隱私期待或是需有隱私的理解的要求。我自己的理解是，個資保護的憲法基礎傾向於理解成人格權下的資訊自覺，而非資訊隱私，這是我個人基於德國法理解的方式。

接下來會進入基本規則的保護面向，在第五條的部分與過去不同是，在歐盟的立法而言，基本規則是直接有效的法規範，指令必須透過轉換的方式達成，與指令不同基本上直接成員國可以適用。因此，在第五條一般個資保護的基本原則，不將之理解僅為為指示性、指引性，應為實體的合法性要件。本身即有拘束個人責任人的基礎，如依照第五條第二項的規定，責任人必須對於遵守第

一項的，負有責任且糸須舉證證明。被認定為，第五條一般原則是界定為實現歐盟運作條約第十六條及基本權利憲章第八條第六項、人權公約的具體化規定，具有獨立性意義在於，當事人在第十五條到第二十二條明文列舉的權利主張外，還可以補充援引第五條主張客觀法效力，違反第五條的情況會導致監督機關措施，以及第八十三條第五項 A 款的罰鍰規定，以下列為六款事由，一一探討會與大數據有衝突之處。

（一） 合法性、公正及透明化

第一款的部分份成三種，這此第一個合法性要件所指為個資處理必須合法為之。於此有廣狹不同的兩種含義，狹義的解讀方式是個資的處理必須要有第六條第一項所定的法律基礎，即必須有當事人同意，或是有第二款至第六款的其他合法是由。廣義的理解是不只是第六條，尚包括所有涉及個資處理所要遵守的義務或要求，如資訊提供義務。多數見解是認為，於此的合法性僅指第六條，因廣義理解欠缺清楚的架構，且相較於基本權利憲章及德國法上的文義都是支持狹義見解。違反資料提供的義務，個資處理並不會因此導致不合法，最多衍生刪除義務、監督措施等問題。故在此將合法性理解成為必須取得第六條法令基礎，如使用德國法用語，第六條規定意指附許可保留的禁止，禁止處理個人資料，例外在第六條的要件下，可以為個資處理。另外，我們將公部門與私部門原則上於基本規則上視作等同處理，我認為這是一個很特殊之處，一般我們會理解基本權在公部門的適用會指向國家性，在私部門的部分會強調，基本權在私人之間至多是間接的第三人作用，惟於此，個資的保護鑒於私部門責任人的地物或是掌握程度不亞於國家，故將公司部門等同處理的構想，是個資法得特徵，也是具有水平效力基本權規範的意涵。第二個公正部分，於德文上是誠實信用，我認為誠實信用的理解上，德國學者認為與民法的誠實信用是不一樣，將之理解為截堵構成要件。當個資處理所受到的不利益，會變成是影響當事人與責任人之間整體地位的平衡關係時，會被認為是違反公正義務的要求。德國學者有特別提到是使用負面表列案例方式掌握有違誠信不公正的行為，如使用隱密的技術、間諜軟體、不比例的個資處理。原則上在此理解下功能較弱，是截堵構成要件為出發點。

在第一款的第三個要素－透明化，所指是個人資料的處理之任何資訊或通知應易於取得、易於理解且應已清楚簡易之語言為之。在此是大數據運用中很大的挑戰，透明化這個部分，大數據的特徵是盡可能沒有目的、廣泛搜集、蒐集的時間長，且能夠利用原先搜集目的外利用更方便，能夠隨之產出價值越高。一開始的搜集目的沒被確定下，而後要透明的告知個資主體就會有困難。在此部分，如何讓個資主體能掌握個人資料如何被處理，是大數據運用上很大的挑戰。此處會認為，越是資料量越大可以創造的價值越多，也會變成結構上越來越難以透明，導致告知後同意的機制很難運作。此情況下，這樣的主張會延伸出的問題是，當事人如請求知悉個人資料是在何種運用法下被掌握、被得出，當事人應如何行使。這樣的主張，如當事人知悉個人資料被大數據運用、分析，

可以找到責任人，要求揭示個人資料如何被操作。另外一方面，在私部門而言會認為演算法、科技設計有營業秘密、投資保障，在德國法上強調會有企業秘密的問題，在此初步的解決方式，德國學者建議引進秘密審理程序，兼顧個資法的審查與透明性的要求，這是第一個基本原則上的討論。

（二） 目的拘束原則

強調個資只能在特定、清楚且合法目的下始得被蒐集，且不得為與此目的不相容的方式再行處理。此具有直接效力，在比例原則上，在此目的匡限出的必要範圍內，才能為個資處理，即手段目的關係，資料最小原則即屬下位概念。此一目的必須於蒐集個資之時點即已確定，且不侷限於蒐集階段，而是直到目的實現為止。目的既已實現，原則上即可刪除，此即儲存限制性原則。從條文觀之目的變更原則上不得為之，目的外利用僅得於相容性審查下，始得為之。且責任人必須負有告知義務，俾使個資主體行使權利。

大數據在此也會有很大的挑戰，基本上我們會希望大數據的搜集是目的開放的，目的開放的情況下，能夠保存越久，日後可能想像的處理手段、演算法的程式能夠分析出的產值也會越高。大數據的分析原本就不希望有目的拘束存在，希望係以結果開放的方式預先儲存資料，亦即不特定之目的。如車輛辨識系統用作辦案工具即屬之。此理解與目的拘束原則是背道而馳的，會導致原本的使用脈絡偏離，以犯罪偵查的例子，會希望以辨識系統為辦案之用，在基本規則中並未有對於大數據運用有明確規範。簡單而言，基本規則對於物聯網、雲端運算、大數據都沒有規定，包括人工智慧。有提到的部分是目的變更的可能性，於此第六條第四項舉出三種目的變更的可能性。此為德國學者在探討目的變更與大數據運用上可能的解決模式。第一是個資主體的同意，第二是特別利用，於此為歐盟法或成員國法的法規。第三是相容性測試，必須與原先目的為相容的情況下。另外於第五條第一項 B 款有提到，基於公益所為之檔案目的、學術或歷史研究目的及統計目的，符合個資的目的拘束原則，也有學者認為大數據運用是符合此的統計目的之用。我個人持懷疑的看法，是否所有的大數據利用都可以被認為是統計目的利用，統計與大數據的分析有些差異，統計基本上是從既有的規則中找出，惟大數據運用不限於要從既有的規則中找出關聯性。

（三） 資料最小原則

此原則與德國法上所說的資料節約或是資料避免原則是相仿的，此強調個人資料處理必須是適當、明顯並且限於處理目的所必要者。此原則是目的拘束原則的下位概念，另一方面，需收集時即為確定收集目的，同時於目的實現時，原則上即可被刪除，即處理目的限制了處理權限。目的本身其不僅要求處理之資料數量必須減少，同時也及於資料使用之數量；資料節約原則也適用於，所涉及之當事人數量應儘可能地降低。亦即「資料禁欲」。這部分與大數據資料運用也有相當程度的衝突，解決之道於德國法上希望透過技術上的資料保護措施，如去連結化（假名化）與匿名化措施，透過第 25 條及第 32 條的立法要求，

回應大數據運用在資料最小原則的挑戰。

（四） 正確性

當事人的個人資料必須實體上正確且必要時，須為最新的狀態；並可採取一切適當措施，就其資料處理之目的而言屬不正確時，可立即刪除或更正資料。這樣的正確性原則課予責任人義務，即必須要積極檢視正確性。責任人透過大數據分析或預測，剖析個人時，必須採取適當之技術或組織的保護措施，修正不正確的個資，並降低此等錯誤的風險，也會連結到第 25 條及第 30 條的部分。還包括在個資主體可於知悉的前提下，拒絕責任人或受託資料處理人所為之自動化決定（基本規則第 22 條）、更正或刪除資料。

（五） 儲存限制原則

此原則指的是時間面向，只要目的達成時，原則上就應該解消。得以識別個人之個資只能就其處理之目的而言，有必要之時間內，始得儲存。依此，儲存限制原則係目的拘束原則與比例原則在時間面向上的具體化。技術上的資料保護措施：去連結化(假名化)與匿名化；第 25 條及第 32 條的立法要求。

（六） 完整性與秘密性

必須透過適當之技術與組織之措施予以處理，以保障免於未經授權或不合法之處理、以防止非預期的洩漏、破壞或損害。此即「資料安全」之一般形式。除基本規則第 25 條及第 32 條外，尚及於第 33 條「個資侵害時通報監督機關」之規定及第 34 條「個資侵害時通知資料主體」之規定。

三、大數據分析中的合法性事由

合法性要件較特殊的是，與個資法不同，不區分搜集的是第 15 條或是第 19 條，或是目的外利用是在第 16 條或是第 20 條的方式。基本上在歐盟法上，僅在處理的概念下，沒有區分不同階段。即該合法性要件不只是處理時須遵守，要在整個處理的過程都需遵守。第二點是與我國法不同的是，並未區分公務機關或非公務機關，原則上皆須遵守，在具體要件設計上會有些差別。第三點是沒有區分不同階段，只是在處理的概念中。在第六條的合法性要件外，德國學者的討論上是窮盡且列舉的合法性要件，除了第九條特別針對敏感性個資，原則上合法性要件都是回到第六條。故第六條的要件設計上，與一般針對各種不同個資處理的態樣分別規定不同，使用概括條款試圖作為合法性要件，立法方式上是沒有區分特殊用途，這是一個很大的不同之處。

問題與討論（敬稱省略）

陳弘儒：

窮盡且列舉的意思是，除了該條所提及的六款外，不能再從其他方式導出合法性要件？

張志偉：

是，當然排除敏感性個資還有第九條。就我的理解，在歐盟法下整個處理階段都必須符合第六條規定。

續上討論

在法釋義學上還是可以找出兩個不同之處，因一個為公部門、一個為私部門，故在公部門部分受到的是禁止原則的要求，如要援引第 C 款及第 E 款，需有內國法的法源基礎，後面會再提及。第六款的一般利益衡量要件公部門也無法援引，可以看出有些差別。故在合法性要件適用各個階段，較不同的是第二款至第六款的合法性要件必須符合必要性原則，即符合比例原則，同意即無必要性原則的規定。另外，強調合法性事由是等價的，並不位階關係。即無需先取得當事人同意為優先的想法。最後一款的一般利益衡量，也並非為最後手段的規定，對於企業而言較有可能使用的是第一款、第二款、第六款，這部分可以分別觀之。

（一） 同意

第一款是同意，當事人同意是資訊自決最典型的展現，當事人同意必須建立在，當事人已經獲知許用目的與責任人的前提之下。同時，被告知的時點必須是於同意前。另外，需符合自願性要求，必須符合當事人自主意識，在個案中受充分告知下，明確的指示下自願為之始屬之。故單純的沈默、預測選項為同意或不為表示，都不是同意。如有選項預設即為同意，必須要點選不同意，這是不行的。必須要明白表示同意，在個資法上才會被認為符合同意要件。原則上無時間上限制，同意後利用個資都是有效的，除非撤回同意。撤回後合法性要件即不存在，即不可為後續的資料處理。故目的達成或是目的解消時，是向未來失其合法性，另外於私部門常見的是，如同個資同意條款，如保險中的各種條約、用途，一長串條約內容，最後需點選同意。在這種定型化契約約款的個資同意上，會伴隨不同意即不給付，德國法上會認為可以援引定型化契約約款的審查，我國民法上也有類似規定，透過定型化契約約款的審查，確保同意性。在德國電信法中有提到連結禁止，如當事人沒有其他管道獲得等價的契約給付或其他管道對於當事人是不具期待可能性，就不能將給付提供與個資同意連結，在德國電信法上是有特別規定的。

問題與討論（敬稱省略）

陳為政：

當事人可以撤回同意，如果撤回同意即需將資料消除。如以人工智慧訓練一

個模型，使用該當事人的資料訓練，當事人撤回，但我已經使用其資料訓練完畢，但是訓練出的參數仍在，是要重新訓練該模型嗎？還是可以保留訓練出的參數？

張志偉：

就我的理解，如可將個人連結消除，也就是你所得出的資料處理結果，不見得是可以還原、連結到特定個人，我認為即以達成刪除目的。也就是，我將同意撤回，故可以請求刪除的是原本個資，惟藉由此個操作出的成果，假使只要能夠刪除，個資成果無法再回復至特定個人，我認為這樣即可。至於技術上如何做，我沒有把握。

陳為政：

我的想像是，不太可能刪除後重新運作一個模型。

張志偉：

如果只是特定個人刪除，如撤回同意。

陳為政：

訓練完成後，我可以刪除你的個資，往後無法再使用資料訓練未來的模型。

張志偉：

我好奇的是，已經訓練出的模型還可以連結到特定個人嗎？

陳為政：

通常狀況下，是不太可能。

張志偉：

如果我要求要刪除原本儲存的資料？

陳為政：

原始資料可以刪除。

張志偉：

原始資料刪除的情況下，還有辦法從人工智慧處理結果得出原始資料是誰嗎？如不行，其實已經達成撤回同意要做的事。

何之行：

就我的理解，撤回同意應該是從撤回開始往後的時點，才會發生撤回效果，不會溯及既往。有個問題是如果今天，資料已經使用，做出的結果應該不會有問題。

陳為政：

比方說我使用你的資料，發了一篇論文，從論文是不可能連結到個人，如果你撤回同意我不可能從寫一篇論文吧。

李建良：

我較不清楚技術面問題，以人作為比喻。經過你的同意我檢視你的資料，檢視完資料後我學習到了，事後你撤回同意將資料拿走，惟該資料仍在我腦中。該撤回的意義只有在於，如果我尚未完全消化，撤回後就無法繼續使用。惟已經學到的東西，是沒辦法消除的，還是是需要消除的？即，撤回對於已經訓練出的結果到底有沒有辦法產生影響？

陳弘儒：

如建模完成，撤回同意的人要求將建模使用的資料拿出，這時建模全部的結果都會改變？

陳為政：

會改變。在大數據的情況下，一個人將資料拿走不會有太大的影響，惟如法律的要求是不能使用且是溯及既往的，等於所有的模型都必須重新運作。

陳弘儒：

因建模並非為0與1的過程，設定條件，靠經驗調整，在此過程中如果發現使用我的資料而未經過我的同意，要求除去我的資料，這是可行的嗎？即還在開發的過程中，並未開發完成，而我主張應該我的數據集全部拿出，假設每個人都這樣主張，不知是否可行？

蘇凱平：

我們最近經歷一個類似的事件，我與師大邵軒磊老師合作司法院刑事廳量刑計畫。司法院說因種種原因有幾個案件要抽掉，即我們的模型都已建置完成。具體影響在我們的案例中，依據兩個因素，一個是該類型的犯罪有多少樣本數，如果樣本數小，操作出的結果就會受到較大的影響，因為有些類型只有一千個案子。如果抽到兩個或三個案子，目前看起來影響不大。如果該類型案子只有一、兩百個，受到的影響就會較明顯，但看不出是什麼具體案件的影響。第二個是原本的偏差值有多大，原本所建立出的模型有些是較偏差的，如果抽掉幾個案子後，對於那條軸線的影響不大。如果原本在該線上，樣本數不大，又連續抽掉好幾個樣本在軸線上的，軸線的斜率便會有變化。

陳弘儒：

在法律上是否能這樣主張，如果我要求抽掉，對於你建立模型會有影響，我拿否這樣主張。

張志偉：

這邊的同意當然是限於同意取得個資，很多時候資料搜集不見得是同意取得的合法性要件，可能是透過其他第二款到第六款的事由，如果是這樣，與此沒有相關。一定是基於個資主體的同意下所取得的個資，事後撤回同意，事後向未來生效。我認為，已經操作出的結果理論上，在法律上不會有法律效果。對於大數據操作出的結果是否會有影響，我認為是運算上需要考慮的問題，這並非法律部門要解決的問題。

續上報告

（二） 契約和前契約措施

債之關係下所必要的個人資料處裡，原則上我們在歐盟法上也會認為這樣是符合合法性要件。因爲締約或是締約前，類似前行為義務的個資資料處理，也會當作是符合的，在此包括所有與契約衍生出的債之關係，包括主給付義務、

服附隨義務，可能推展的進程也算。在學理上較有討論的部分是，如果只是單方的債之關係，如懸賞廣告、中獎獲利，或是民法上好意施惠關係，在這種情況下是否能使用這款事由，容有討論之處。如不去深究細節，基本上這款事由要求必須要有必要性，必須符合、貼近契約目的以實現為必要，即個資處理與債之關係有實質聯繫。契約的給付或是內容必須仰賴一定資料處理時，才可以收集個資。這款事由我們會認為在古典契約樣態上不太會有爭議，惟新型態的服務是長期性，如智慧型電視，是長期個資處理問題，此部分會較難解釋是否符合這款事由，在此也會要求必要性原則必須更嚴格，因為使用時間是長期。

（三） 法定義務

第三款是法定義務，責任人依法所旅行的法定義務亦可作為資料處理之合法性要件。在此非獨立合法性要件，必須有歐盟法或是其他成員國的法律基礎。這款在學者間的討論會認為有指令性質，必須透過其他法規範連結，可以想像在德國法連結上，包括像是交通許可或最低工資法。這款也要求必要性，法定義務的範圍內才可以為個資使用。

（四） 生死攸關之利益

第四款是生死攸關之利益，如果看萬國翻譯是翻成重大利益。惟視德國的討論，照字意翻譯是生死攸關或對於生命重要的利益。參其舉例態樣，大部分都是與個人生命、身體有密切相關，我會較傾向翻成生死攸關，較凸顯出這款的特徵。這款事由較重要與天災人禍有密切相關，在這種情況下的個資索取，當事人不見得有能力同意。援用這款的情況，當事人可能已經無法表達同意，此要件有點類似於推測同意的態樣。限制上，有學者認爲援用必須是非敏感性個資，且當事人無能力行使自主決定權，即無其他合法性要件，才可使用這款事由，必須要符合推測同意內涵。

（五） 公益或執行公權力

第五款是公益或執行公權力，這款是公部門資料處理的核心，與第三款類似，即必須有具體的歐盟法或是成員國法才能連結此款作為合法性要件，故此款也會有指令性質的特徵。在此重點是公務的履行，因此這款事由並沒有要求供行政機關，如果是私人處於委託行使公權力的情況下也可以援用此款。重點是內容，而非身份別，包括職業工會、私法人都可以援用此款作為要件。在德國法上的政黨、宗教團體也是可援用的要件。

（六） 責任人或第三人之正當利益

第六款是概括事由，指涉的是必須為利益衡量。個資處理是必須為資料人的正當利益所必要的情況下，只要個資利益沒有優先於責任人或第三人正當利益，就可以援用。是非常廣的，德國學理上會認為是核心的衡量條款。優點是彈性，因為指示操作利益衡量。而缺點是在於優點，因為太彈性。第六款的這些要件底下，第六款是窮盡列舉的性質，不可避免地會有概括條款，否則無前面幾款的合法性事由時，就要放棄個資運用，可能會對於個資資料流通有疑慮。故在此款會要求必須做私人利益與私人利益的衡量，為何沒強調公私利益，因

機關履行公任務是無法援用此款，是典型的古典衡量法上的要件。限制是不能是機關在執行行政任務所為的資料處理，如機關是以私人主體的身份參與法律事務時，可以適用，如行政法上私經濟行政，即可援用。不拘泥於身份類別，而是視其任務型態。

四、大數據應用在個資法上的總體評價

在此的立法模式不是針對特定情境或是特別領域資料處理之實質合法性規定，是以抽象方式，所謂一體的方法適用合法性要件的立法模式。故會造成法律適用者的挑戰，德國學者也有批評，對於法適用機關是很大的挑戰，因伴隨很大的裁罰，解釋上的不安定性應盡可能排除。個資的監理機關必須提出對於個資立法上的解釋，類似解釋性行政規則的理解方式。在大數據的運用上，也強調必須遵守一般的個資保護的法規與具體個別的特別法。故也適用在個資法上的一般禁止原則，即原則上禁止蒐集、處理、利用和儲存個人資料。如個別領域有大數據運用的需求時，可以另訂合法性要件。在德國學者介紹上，認為很多個吃討論的熱門議題沒有處理，是較簡化的立法，是第五條與第六條所遇到的質疑。我認為是，但立法模式並非用在特別情境，可以想像在德國個資法或反觀台灣個資法，也是類似情形，即使用抽象構成要件體現合法性要求。

適用上可能面臨的問題在於，與基本權所衍伸的衝突，如目的拘束、資料最小、資料匿名、資料節約與透明原則等扞格之處。故會導致，如使用同意作為合法性要件，目的外變更時，很難再找到當事人在要求重新取得同意。另外，有規範大數據運用是在第 89 條提到，基於公益的檔案目的、學術研究、歷史研究或是有統計目的的例外條款，會涉及較簡易的認定目的外利用、目的變更情況，或是儲存時間會變長，或是請求刪除的權利會加以限制。這是基於學術研究、檔案目的或統計目的的特性，而給予的特權或優惠。類似疑慮在我國個資法也會有同樣問題，大數據運用時責任人無法履行告知義務，同時搜集目的消失後不得持有、處理、利用的規定，很難想像與大數據連結。另外，特定目的運用上很難配合，很難為拒絕權，其他當事人權利在大數據運用上也很難貫徹。

參、可能的解方

一、匿名化或假名化作為解方

這裡會有的疑慮是，是否能透過匿名化、假名化的方式做，取決於對於這些概念的掌握。在法條中都是以「無識別可能性」、「無從識別」為規範，惟「無從識別」觀法務部的相關資料，其實是大數據運用提供給其他公務機關或是學術研究機關，重點是研究成果發表時依其呈現或揭露方式無從識別。內部的傳輸、交流或是傳遞給私部門，都沒有要求須匿名化，而是在最後的階段成果發表時才要求要阻斷識別可能性。這樣的理解會與匿名化要求差很多，法務部只要求成果發表時無從識別即可，頂多是去連結化的假名化措施，並沒有達到匿名化的要求。我認為國內在討論去識別化、匿名、去連結、編碼、加密、假名

化，這些概念每一個用語的使用是不夠精確。有個學者提到去連結就理解成為匿名資料，對照前面提及個人資料的理解，個人資料的相對概念是匿名資料，也就是匿名資料是沒有適用 GDPR 的。如只是連結或代碼，惟既沒有分開儲存，也沒有同一個當事人間，同一個責任人是否可同時持有連結或代碼符號，這甚至連假名化都達不到，我們似乎認為將之圈選之即為匿名資料。我認為並非正確，但我們也沒有照敏感程度區分資料層級、不同的保障程度、去識別化技術設計不同層次，這我認為也是有疑慮的。實務上對於匿名化、假名化的使用非常腐爛，但對其實沒有定義的，這是很大的問題，我對於第一個解方是有疑慮的。

問題與討論（敬稱省略）

何之行：

我可以理解去識別、假名化完全沒有達到匿名化的程度，關於去連結在台灣法律的規定中，有人會認為去連結就已經像是歐盟定義下。去連結與去識別的標準在於去連結必須做到已經沒辦法直接或間接使用，人體生物資料庫管理條例是說已非連結的方式保存。

張志偉：

只要還有連結可能性我們會認為還是不足夠。

何之行：

如果是已經將資料處理到所謂去連結，他們有共識是這樣的資料有點像是匿名化資料。

張志偉：

但是他的回覆可能性有多高？在連結化的可能性有多高。

何之行：

台灣規範不像美國 HIPAA 那麼清楚，不管技術上做不做得到，規範就認為是，也未清楚規範假名化的個資還是個資。我們是使用去識別化，卻不知道如何做。

吳全峰：

我認為是術語問題，因為美國使用 deidentification，可是美國的 deidentification 與之行說的一樣，HIPAA 是十八個，可是不同的法律下規定，去除的變量不一樣多，故他們使用 deidentification。可是他們另外有一個術語是 delink，台灣翻譯成去連結，但 delink 在美國的定義是永久無從連結，但是有「永久」，類似台灣人體生物資料庫管理條例的規定是永久無法回復的狀態。假設照該定義理論上穢語 anonymization 是類似的，惟歐盟的 anonymization 也並未走到如此極端，因為歐盟的 anonymization 不認為有永久無法識別這件事。故歐盟的 anonymization 認為是時間、資源允許的情況下，是無法再回去識別。台灣的狀況是人體生物資料庫管理條例有界定 delink，可是在細則中的 delink 又認為編碼可以達成 delink 的效果。在實務操作上常使用編碼，就認為達成 delink 的效果，與之行所說的一樣，衛福部針對此是沒有任何的解釋的，當一個研究者拿一個編碼問是否有達成 delink 的效果，我們無法認為沒有，問題在這。

何之行：

不只翻譯問題，每個用法都不一樣，我們沒在規範上像是 HIPAA 或是 GDPR 給予一個清楚的線，或是技術上可以做到，但是規範上定義，我們都沒有這樣做。

劉靜怡：

我會認為有些基本規定，故接下來的施行細節或是指導原則本來就有很好的機會，將這件事釐清，但是主管機關故意混淆，問題是在這。

吳全峰：

以個資法的規定而言，是無從直接識別或是間接識別當事人，這是法律的規定。如無從直接或間接識別當事人，事實上與 GDPR 的規定是非常類似的，GDPR 的 anonymization 的定義也是無從識別當事人。可是 GDPR 下並未給出如何達成的方法，HIPAA 的十八個也有可能是這種方式。到了其他法律時，好像變成只有 HIPAA 才是去識別。狀況是，HIPAA 的去除十八的 identifiers 不一定在所有狀況下都是無從直接或間接識別當事人。假設是資料量夠大時，說不定十八個是不夠的。故歐盟是一個概念匡列無從直接或間接識別當事人，我從其他方式檢驗是否可達成該要求。美國的去識別化較類似給予一個標準，如達成該標準及回 deidenfication。台灣的概念混淆，在個資法規定無從直接或間接識別當事人，要達成時又將如十八個方法套入，而未給予風險值。我會認為台灣的狀態是硬是湊在一起導致這個縫隙。

李建良：

我有個方法論的問題，前面都提及 GDPR，現在進入至可能的解方。可能的解方的第一個問題是匿名化或假名化，這個的問題設定是放在歐盟的法秩序下，再回到台灣的法秩序，可是從第一個問題就切入至我國的個資法。所謂術語的問題基本上是涉及特定法秩序，在該法秩序下理解後再去評價，做為問題解決是否能達到目的。可能在順序上談匿名化或假名化在歐盟的 GDPR 體系中，是否為一個解方。我們的規定是在 GDPR 之前所訂定，故使用 GDPR 的規定看我國的個資法，只能夠為遵循或是繼受。

劉靜怡：

較合理的比較點是，比較 1996 年的 EUW。我們在第二個版本的個資法，嘗試想要模擬該精神，事實上立法者修改，加上行政機關施行細則與解釋將之扭曲。

李建良：

在 GDPR 的法秩序下，還是一個問題。即二分，如是匿名化或假名化不在 GDPR 的適用範圍，將之切斷，惟切得沒辦法完全斷。不是百分之百完全不可能，這個規定在他們法秩序下可能就是一個問題。在他們的法秩序下不會是解方的情況下，我們要引進或做對照時，是否有可能做得比他們更好，如果我們繼受他們規範，可能也會將問題繼受。如果真的要比較這個部分，應比較 1996 年，惟 1996 年與現今 GDPR 又有進展。

劉靜怡：

故較公平的比法，將現行的個資法比 1996 年的 EUW，個資法修正草案拿來與 GDPR 比較。

林勤富：

現在看到很多的名字，其實有些是法律規定的結果，有些詞如編碼、加密是達到某種結果的方法，有可能達不到該結果。現在將之混用，施行細則將用了這個手段，是為符合該結果，會有錯誤的配對，是因為法規制定包含 GDPR，其實都不了解技術的極限或是技術了厲害。如將結果規定在法律中，永遠沒辦法達到該結果，因永遠都有辦法回溯識別該人。如法律不管技術，要規定該結果是永遠無法識別該人，該結果是沒有用的。除非如同美國使用手段，符合一定

程度，對他們而言即為符合 HIPAA 規定。會變成，第一不能將手段與結果混合，否則將會產生現在這種情形。另外是從技術層面，將結果、想要達成狀態訂定在法律當中，也許在技術上是不可行或是沒有用，如果還是要將結果訂定在法律中，執行上會有問題。

劉靜怡：

可以想像為何會這樣做，立法者說不得直接或間接識別個人，目標是這樣，無論實際上是否得達成。於是，行政機關在編施行細則時，開始自行想像、拼湊，如果這樣就已經等於不可以直接或間接識別，但這並不是。所謂不可以直接會間接識別個人，基本上可能是要維護資訊自主權，即無法識別個人，自主權就有一定的保障程度。施行細則、函示的內容，其實是行政機關自行填充後的結果，行政機關不知道可以找到何種東西是百分之百無法識別出的。除非行政機關自行承認技術有限，但他們並不會這樣做。又被逼著必須有那些不可及的技術，可能在立法架構下就是要去修改，因為這樣的架構必然會產生問題。反而大家在細節或是末端想盡辦法要解開結的結，結果是往上掏空搶想要追求個資保護的目標，函示的結果就會是，所以東西都可以不是個資。

何之行：

有個規範上的問題是，美國 HIPAA 的好處是規範上定義很清楚。但他其實在大數據時代開始後被批評得十分厲害，因為完全無法處理 reidentify，而且有很多案例發現 HIPAA 很容易被 reidentify。立法上應追求規範上讓實務可以很清楚操作，還是應反映技術是一直在革新。歐盟提出隱私風險評估，反而是希望是透過此處理這個問題。

李建良

就上述討論可以得知，這可能不是一個好的解方，不管是哪個國家或是法體系。

吳全峰：

後來美國與歐盟事實上並沒有差得太多是因為，以美國 HIPAA 而言，是有決定空間的。不是 HIPAA 十八個手段是死的，而後還要風險評估。美國專家當初決定該十八個時，是有做過風險評估的。做完的問題是沒辦法選「值」，故後來選出十八個是某種程度在不得已的情形下，故是有風險評估的。歐盟在匿名化的部分，實際上是假設某工具在風險評估上可以達成某效果，即可以認同該方式是可以的。又以編碼而言，台灣的編碼以身分證字號編碼，基本上很容

易被編碼。惟如編碼加上後面二十個選項後，編碼又可能容易被破解。編碼是工具，不是結論。去連結在人體生物資料庫管理條例中規定，在細則中提到去連結可以以編碼方式達成，理論上是以編碼方式達成去連結的法律定義，但在操作上變成，既然想要編碼，即編碼，但不會再檢視與法律本身去連結定義的連結。

劉靜怡：

施行細則於我而言即已超出母法的授權。

張志偉：

歐盟法在法條上，本身只有匿名與假名，甚至匿名根本沒寫，GDPR 中找不到匿名這個字。但學理上、實務上都認為匿名資料就不是個資，所以只有規定假名化，假名化方式可以編碼、加密等達成，這個部分也沒有這麼多概念在條文中。令一點是，強調這樣的假名化方式使用代碼或編碼操作，應是要分開保管，組織上保存技術，實際上情況運作如何我不清楚，惟我國連分開保管都沒有要求，這是我認為達成的風險本身是滿高的。

續上討論

二、透過科技形塑以及對資料保護友善之預先措施

在歐盟法上有很多討論的是，透過預設著手保護，透過設計著手保護，及在前階段就進行保障，要求義務人要採取適當技術性及組織性措施確保。實際實務上是否能達成是另一件事，這是我認為應努力的方向。

三、由資料負責人與監督機關負起個資保護責任

第三是要求由資料負責人與監督機關負起個資保護責任，要求公司具有一定規模是有資料保護專員，保護的後果影響評估、定型化約款的審查。我國學者有提到，在演算法結構上、認定上要討論，演算法的要求是否有符合個資法，原則在監督機關的理解上，只要事後控制即可，有特別提到針對敏感性個資演算法的審查，認為需事前審查。最後驗證與驗證機構制度，我看到我國翻譯上都翻譯成認證，如果看原文，將像是公私部門協力的機制，故我認為應以驗整理解。

四、當事人權利的強化

強調在基本規則第 22 條有自動化決定的自決權，此部分是在歐盟法上談論大數據、演算法上很重要的一環。要求如個人資料是被自動化處理而做成，對於當事人具有法律效果或是顯著影響時，可以有拒絕權。拒絕權實際上如何操作，是下一步我想要研究的部分。當事人之資訊請求權、更正權、

刪除權、限制處理權、資料可攜權，都是歐盟法下想要強調個人自主的空間。

肆、結論

我認為個資法應配合修正，因之前參考的對象並非現今的GDPR，要爭許適足性、認證時，我認為有再修法必要性。另一方面，專業法規上的授權條款也是必要的，個資保護不會只有在個資法中達成，需要其他專業法規的授權條款，更嚴密、詳細的合法性要件的審查，這部分我認為是可以降低大數據影響下的個資侵害的風險。

第六次會議紀錄	
時間	109 年 12 月 18 日（星期五）
主題	以人工智慧輔佐法院心證？統計證據的觀點
講者	蘇凱平（國立台灣大學法律系助理教授）

內容摘要

壹、　延續：AI與刑事審判

在刑事審判的進行過程中，從認識法律開始、到認定事實以及判斷證據之後，最後要做的事即為量刑。而從刑法第57條來看，我們可以看到有許多量刑因子，就此可以說是審酌一切的情狀，也因而導致每一位法官在量刑時可能差距巨大。而這也是目前司法院在推動國民法官法遇到的巨大問題。對於職業法官而言，在受訓的過程中，事實上有一定的標準；然而，對於國民法官而言，他們並沒有這樣的基礎。

而一般人即認為，關於量刑的議題，相較於對於證據的判斷，AI的輔助似乎就有較高的可能性。然而，可能要注意的是，事實上在證據的類別中，有一類型是為「統計證據」。統計證據是一種數學化的證據，而就此是否即可以透過AI的輔助更幫助法官達成心證？因為就目前AI可以達到的技術而言，主要有整理事實的能力、抽象涵攝的能力以及計算的能力；AI比較受到質疑而不具有的能力是說明理由的能力，然而在與多制度諸如美國陪審團制之下，說明理由並非必要。因此，AI事實上在刑事審判中的過程應能妥善應用。

所以更關鍵的問題在於，究竟法院裁判有何要素為AI所不可取代？在關於證據法的討論之中，普遍認為人類法官在運用證據判斷事實的過程中，是一個模糊的狀態，是一個難以測量的洞見（ambiguous implications）。此外，人類法官可能有情緒控管的問題，而且也可能帶有有意識或無意識的歧視或偏見。而對此而言，機器人法官會不會比人類法官做的更好？

貳、　框架：AI與統計數字運用

進步言之，本文的討論層次在於，AI如果技術上能達到這樣的任務，我們要不要讓它這麼做。亦即，本文討論聚焦於AI的「應與不應」而非「能與不能」。以COMPAS為例，常受到的批評是它有種族歧視的問題。然而，COMPAS在做的事情有三：之一是關於對是否應予羈押以及其他替代措施的評估，之二是評估再犯分數有多高，之三是對於暴力犯罪的統計。其中，關於再犯分數的統計，其計算方式為：將年齡、初次被逮捕的年齡、過去暴力犯罪的歷史、正當職業的可能、過去不服從法院命令的歷史等因子分別乘以權重。而這樣的計算的方式，究竟歧視的問題存在哪裡？關鍵在於這樣計算出來的結果可能會有偽陽性（false positive）的問題。亦即，機器可能預測出，黑人實際上被認為不會再犯，但卻被算出較高的再犯分數。然而，仍有幾個州再繼續使用。原因在於，這些州認為，這樣的計算方式仍是針對個人，而不是針對個人所代表的群體。

陳弘儒：如果是針對個人，那這樣的系統是如何畫出權重？

蘇凱平：對此COMPAS公司有手冊，告知權重如何被運算而得出。

續上報告

因此，要去思考的是，所謂的偏見與歧視究竟所指為何？就此須與以區別的是，主觀性（subjectivity）的偏見以及客觀化（objectivity）的偏見。前者是指，人類的判斷者會有主觀的偏好，而可能帶有人類的情緒，因而有主觀性的偏見。而後者則係指，這樣的偏見是來自於被判斷者。亦即，透過個體所述群體的行

為，去判斷這個人，而由此可能產生的偏見。比如因為你是個黑人，住在特定區域，因而有負面的評價。而AI會不會就是有客觀化偏見的問題？因為其必須用其他東西去衡量、計算以及模擬，而不是針對個體的具體情況。

此外，類似的討論還有關於美國的信用評等系統對於個人信用分數的計算，是否有歧視的問題？而一般人會認為沒有，理由在於，關於信用分數的計算還是針對個人的信用歷史，比如是否超借現金、是否逾期還款、是否未繳款等。而這些因子都還是個人所導致（How have you behaved in the past?），而不是以跟個人有關的群體去衡量計算（How have people like you behaved in the past?）。因而被認為沒有歧視的問題。

因此，要去思考的即為，後者這種以跟個人有關的群體去計算評估的方式究竟是否可行？這種統計學上通常能成立的模型，是否容許被應用？又如何去面對被錯誤分類，造成波及損害（Collateral Damage）的個人或個案？比如一個人可能只是因為剛出社會很窮因而住在較差的區域，由此去推得其容易再犯、容易脫逃。這樣是否要容許法院在審判中進行運用以及考慮？刑事審判中應該要如何去處理這樣的問題？

參、　現況：刑事審判中的統計證據

在美國法的討論中，禁止運用統計證據直接去導出無罪或有罪的結論，而只能以輔助周遭事實的認定。以People v. Collins, 68 Cal. 2d 319 (1968)案例為例。案件事實為一對夫婦被指控為強盜案嫌犯，被告夫婦的外型特徵包括：丈夫為黑人，有時下頷蓄鬍；妻子為金髮白人。而目擊證人即在審判中證述看到這樣的嫌犯特徵。審判中，檢方的主要任務是說服法院：被告確實即為搶匪。檢方以統計專家做為專家證人，提出「搶匪恰好與本案被告具有相同的上述種族與外型特徵之可能性」意見，認為僅有一千兩百萬分之一。比如檢察官主張金髮女性又綁馬尾的機率為十分之一、黑人男性留有落腮鬍的機率為十分之一、白人女性與黑人男性結婚的為百分之一等；而把這些獨立事件的機率乘起來，即為

這件事會發生的機率。而法院最後即判決有罪。

後經上訴到加州最高法院而被駁回。加州最高法院提出的質疑有三：第一，這樣的判斷欠缺實證的基礎；第二，這幾件事情並非獨立變項，計算方式即為錯誤；第三，這樣的統計證據使用即會誤導陪審團。就此，即能說明對於統計證據的運用即涉及關於數學的計算，而可能成為數學帶來的毀滅。

而回到我國刑事法院中使用的統計證據。首先先看高院的這個判決（台灣高等法院102年度上易字第2280號刑事判決）:「實務上常針對DNA-STR型別實施檢測，將檢體之細胞核DNA進行分析，取得15組STR數值與性別染色體，再以統計推論所得之特定人口中DNA型別重複出現頻率為基礎，計算15組STR均相同之機率，如該機率數值甚微，代表該特定人口中幾無可能出現另一相同DNA-STR型別之人…足徵被告確曾至遭竊現場飲用上開寶特瓶所盛裝之飲料，是被告前開辯解，並不足採。」而事實上這樣的關於DNA的統計證據的使用方式，在我國與DNA有關的判決中，法院都會這樣主張。

再來是關於血中毒品濃度的判斷，關於毒品濃度多少才會導致不能安全駕駛在法律中並未有所規定，而在最高法院的這個判決中（最高法院107年度台上字第205號刑事判決）即表示：「況所謂不能安全駕駛，非係以瀕死亡、休克為判斷標準，法醫研究所經參考各國統計分析結果，認定血中甲基安非他命在500ng/ml以上，即構成不能安全駕駛，係依一般客觀情形判斷，常人若施用毒品達上開尿液濃度，已達不能安全駕駛之狀態，上訴人既已符合上開要件，應認已有不能安全駕駛之情。」由此，我們也可以看出，法院認定的方式是透過「其他人的情況」來判斷衡量在使用毒品的情況個人是否不能安全駕駛，而不是該個人的情況。

此外，尚有關於醫療糾紛的案件。在這個判決（最高法院102年度台上字第809號刑事判決 ）中，最高法院表示：「又脂肪栓塞除了造成肺部血管之栓塞外，全身其他器官血管是否都會有可能發生栓塞？腦部發生機率是否最高？…統計上，在沒有被考慮『脂肪性肺栓塞』診斷的病人中其死亡率多少？…敗血症、

脂肪栓塞、血管內凝血溶血症（DIC）、肺部感染後導致之敗血症，是否均為引起急性呼吸窘迫症候群的原因？統計上，其死亡率若干？」亦即，其認為前審未考慮脂肪性肺栓塞的死亡率是多少以及其所引發而導致之後的敗血症在死亡率上又是多少？然而，問題在於，法院就算考慮了這些統計數據，就能在個案中去認定醫生的操作是有過失的嗎？

再來，在證券交易法領域也有關於統計證據的使用，關於內線交易的消息重大性又應如何認定？最高法院即有認為應將事件發生的或然率（可能性）列入考慮。

肆、　討論：比較的對象

因此想提出討論的是，對於統計證據的使用，加入AI的判斷來輔助會不會比較好？而這又可以分為兩個層次來討論。第一，使用統計證據本身是否即為一種偏見？而這有幾個可能的回答選項：認為是偏見，因為是利用個體所屬的群體進行判斷；也可以認為不是偏見，因為證據的性質或者證明的本質就是如此。而第二個層次的問題是，法院審判中，原本即容許統計證據的使用，是否也應容許使用AI輔助法院判斷？而這也有三個可能的選項：第一，一概的允許AI進入輔助；第二，在原本容許使用統計證據的類型，讓AI可以進入輔助；第三，一概的不允許。

問題與討論（敬稱省略）

李建良：

在People v. Collins這個案件中，一審陪審團的認定不需要理由，而在上級審由法院認定。上訴審如何去認定一審陪審團的判斷的理由是什麼？因此想問的是制度上的問題。

蘇凱平：

在美國的審判制度下，對於事實審的審判是不附理由的。在此之下，上訴審不會去認定一審判決理由錯誤，而是看「上訴理由」是否有道理，也就是去看指摘一審錯誤的理由。

黃詩淳：

這邊討論的證據的意義是什麼？我們在討論的統計的東西比較像是經驗法則的問題，那在美國法上是證據的問題嗎？那要怎麼去思考這樣的證據能力的問題？

蘇凱平：

關於證據能力的討論是重要的問題，也就是是否能將這樣的證據呈現在陪審團前。有時候會有premier hearing，由專業法官先判斷證據有沒有證據適格而得以做為證據；有時候會是在審判中爭執，由辯護律師和檢察官在審判過程中爭執。但在本案中情況如何並不清楚。

陳弘儒：

這個案子可能的詮釋是，檢察官用的是機率論；然而這個在統計上的問題是，條件機率還有貝氏定理，是比較後期才發展出來的。在這樣的基礎上，再去應用於在既有的證據上，去檢驗對於假設成立的可能性有多高。因此從學科史的發展來看，可能再當初的確是有一個落差存在。

另外一個討論是，在COMPAS的情形，我會很好奇為何法院最終會認為算出來的數字是個別。也就是說在數學運算上，關於權重的設計並非針對個人因子，而是群體結構所得出的係數。還有一個問題是毒品案件的那個問題，我認為法院在此並不是在使用Proxie，因為那還是國外的數據。

蘇凱平：

關於COMPAS公式中的權重，我也不認為這是完全個人化。在它的手冊當中，它就寫這是經過研究之後，所得出的數字。所以我們也可以理解為何美國的多數州並沒有採用這個系統。

陳弘儒：

我的意思是，它是拿國外的數據，所以不是proxie概念。因為台灣沒有這部分研究，我們看不到數據。我認為這邊只是把國外研究套到自己身上，而不是到使用替代指標的概念。

邱文聰：

這邊有幾個概念可以再釐清。你這邊似乎把proxie和個人化因素混淆在一起了。zip code是一個proxie，但它還是會指向個人化因素；這跟國外的研究能不能引用，又是另外一個層次的問題。你的比較核心的攻擊點應該是在說，個人化因素不可能是真正的個人化。使其看起來像個人化，但它還是建立在個別的人所屬群體的統計上的特質。

蘇凱平：

在這個意義上，是不是沒有所謂的個人化因素呢？

邱文聰：

所以真正的核心問題是，我們在什麼樣的情況下，可以合理使用統計證據？而這會繫諸於制度設計目的到底是什麼的問題。比如我們要雇傭一個新人，我們就是基於它過去的表現來預測他未來的表現。但他還是基於一個一般化的因

素，也就是相較於一般同儕的表現去比較。所以從這個角度來看，統計證據的使用是有效且符合目的的。而在刑事審判上，也是要看的是統計證據的使用有沒有符合目的。

蘇凱平：

除了按照制度目的來判斷是否合理外，也有人討論的是手段是否合理。也就是具體使用的因子到底是否合理，對於制度目的之達成是否妥適。

林勤富：

我也覺得COMPAS不是個人化的計算方式也沒辦法達到。最高法院也不這樣認為。所以才要加警語說，這是基於群體所得的資料。但從制度設計來看的話，COMPAS建議的風險值不是一個個人化的數值，所以只要法官不要完全使用，然後有參考其他資料的話，他某種程度上就已經達到個人化的考量。

何之行：

我們對AI的要求是希望能夠透明。而關於COMPAS權重因子都已經公開，該當事人就可以做一些行為選擇來改變結果。但在統計證據上，他的透明化以及揭露可能性對於當事人來講是不是很難做一個行為選擇？

蘇凱平：

如果很廣泛地以統計資訊來說，去改變是有可能的；在刑事法比較已經確定而無法改變的事。

楊岳平：

關於COMPAS再犯分數的計算，他的問題不是在於能不能做這樣的預測，而是

他用的有些因素是不合理的因素。比如：年齡、職業教育程度等。關鍵應該是在於用了什麼因子在裡面。

蘇凱平：

而這些是行為改變動機，不一定是壞事。而在此例子中，比如暴力犯罪史，站在法官面前已經不能再被改變。但是像職業教育程度，在死刑無期徒刑的案子中就會盡力去主張，而也有法官真的會去進行考量。

邱文聰：

我記得COMPAS是137個因子，然後現在是濃縮成五個變項。那他跟我們想像的人工智慧的訓練不太一樣。因為這裡是人類挑選五個重要的因素，而AI是他不知道那些因素會影響再犯。所以COMPAS他應該不是用機器學習得到的公式吧？

黃詩淳：

COMPAS現在是用機器學習方式得到的。

蘇凱平：

是。過去沒有機器學習是用分析案件所得到，後面才加機器學習。

邱文聰：

會問這個問題是因為在談機器學習的時候，會說機器學習的東西不需要理論，而是讓它下去跑。而現在這看起來像是理論，用來引導特定行為的改變。應該是不一樣的徑路。

陳弘儒：

我認為應該還是有用機器學習，但問題在於權重的意義如何。透過機器學習的類神經網絡才連結時，GOOGLE也搞不清楚意義是什麼。所以如果權重是透過機器學習，那麼意義為何才是需要被思考的。此外，更大的挑戰在於，過去的資料是否可以精準未來。

蘇凱平：
這邊法院是說不可以只用COMPAS去判斷，所以這是應與不應的問題。而不是說AI在技術上做不到這樣，但剛剛弘儒講的是技術上做不到。所以你應該是覺得如果技術上做得到就可以。

吳建昌：

我比較質疑的是預測模型，我們都是餵資料給AI。所以如果拿30年前的資料訓練出來的AI也無法用於現在的情況。而在預測再犯這件事，有太多事情要做了。而且未來的社會環境情況也一直在改變，我不覺得AI也可以預測未來的這些事情。很多情況都還是用猜的，而我們可以接受多高的失誤率？

蘇凱平：

預測這件事情畢竟是預測，能不能做得到跟應不應這樣做還是兩個問題。

楊岳平：

補充一下FICO的部分。他事實上也沒有把所有可以用於預測的因子都放進來。比如說種族沒有放進來。而些因素之間都是進行權衡。而設計系統的目的絕對不會是單一的，而有其他需要平衡考量的目的。因此哪些目的應該要被考量才是關鍵，這還是需要有人去做價值判斷。

蘇凱平：

最後如果要一個改變行為模式的效應，還是需要手段和目的去配合。

張兆恬：

我國法的四個例子，使用的統計證據，應該是處理不一樣的問題。比如說毒品那個是在處理有無安全駕駛能力；而醫療糾紛的案子，處理的問題是討論機率問題。而這是處理不一樣的事情，所以討論要不要應用統計證據的時候，所要考量的事情是不一樣的。

蘇凱平：

而這些確實都是不一樣的。

高國祐：

想請問老師所討論的統計證據究竟所指為何？因為感覺有的是針對構成要件的解釋，有的又是針對個案中的事實判定問題。前者比如說對於毒品濃度不能安全駕駛的程度，到底多少是不能安全駕駛去進行定義，而這是大前提的問題；後者比如說醫療過失的案子，去個案認定事實有沒有未考量具體的死亡率。不知道這個是在美國法中對於證據會這樣使用嗎？因為前者應該不是用證據去認定個案事實的問題。

蘇凱平：

在英美證據法，證據法是全部都包括的。

林昕璇：

關於COMPAS的應用有沒有發展出可解釋性的AI，去彌補刑事審判實務上可能

受到的透明性與黑箱的批評與挑戰？

蘇凱平：

在美國的案子中是用應然面去做規範，也就是規定不能只看AI所跑出來的結果去判斷。而在我國法上制度上沒有相關的討論。而在我國的量刑資訊系統只說可以參考，但還是要在個案中進行具體的認定。

李建良：

可能要確認的是，既然是AI做的是輔助性地判斷，最終的決定權還是在人，在於法官。法官還是要看到當事人，而這是AI沒有辦法取代的。而在COMPAS的案子，都沒有提到再犯率其實可能跟本案的犯罪內容還是有關，而那個公式並沒有這個因素。此外，統計數字如果要應用還是要回到規範面下。而你舉的例子很多不是證據去認定事實的問題。比如毒品的那個是因為法律沒有定濃度，所以何謂致不能安全駕駛就要去進行判斷，所以這不是證據問題。而是這個濃度要不要被認定為不能安全駕駛，這個是評價的問題。

參考文獻

(一)中文部分

期刊論文

1. 王伯霳(2018)。〈美國聯邦審計署表人工智慧技術評估報告〉,《科技法律透析》,30:8 期,頁 4-5。
2. 王明禮(2014)。〈論資訊隱私:科技與商業發展脈絡下的觀察(Information Privacy: A Contextualized Analysis)〉,《中原財經法學》,32 期,頁 59-105。
3. 甘琳(2018)。〈歐盟執委會公布歐洲人工智慧通告〉,《科技法律透析》,30:6 期,頁 6-7。
4. 何亦婕(2015)。〈日本推動智慧醫療照護與巨量資料應用之趨勢觀察〉,《科技法律透析》,27 卷 12 期,頁 51-69。
5. 吳旻純(2011)。〈人工智慧新革命——超級電腦「華生」〉,《生活科技教育月刊》,44 卷 5 期,頁 1-9。
6. 吳柏凭(2016)。〈人工智慧對於著作權概念的衝擊——日本著作權的新政策發展方向〉,《科技法律透析》,28 卷 12 期,頁 26-31。
7. 宋皇志(2017)。〈人工智能在專利檢索之應用初探〉,《全國律師》,21 卷 10 期,頁 27-37。
8. 林利芝(2018)。〈初探人工智慧的著作權爭議—以「著作人身分」為中心〉,《智慧財產權》,237 期,頁 61-78。
9. 林芝余、陳婷(2018)。〈人工智慧與雲端運算法律相關議題〉,《理律法律雜誌雙月刊》,107 年 1 月號,頁 4-5。
10. 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。
11. 張保生(2001)。〈人工智能法律系統的法理學思考〉,《法學評論》,19 卷 5 期,頁 11-21。
12. 張麗卿(2019)。〈人工智慧時代的刑法挑戰與對應──以自動駕駛車為例〉,《月旦法學雜誌》,286 期,頁 87-103。
13. 曹源(2016)。〈人工智能創作物獲得版權保護的合理性〉,《科技與法律》,2016 卷 3 期,頁 488-508。
14. 郭雨嵐、汪家倩、侯春岑(2017)。〈法律科技與人工智慧時代,科技法律人才的養成與挑戰〉,《萬國法律》,214 期,頁 51-59。
15. 陳良基(2017)。〈打造人工智慧創新環境機制〉,《國土及公共治理》,5 卷 4 期,頁 60-71。
16. 陳譽文(2017)。〈人工智慧規範性議題綜觀〉,《科技法律透析》,29 卷 4 期,頁 43-51 頁。
17. 黃崧洺(2018)。〈從人工智慧與法律科技展望法律服務之價值——以專利

服務為例〉,《專利師》,35 期,頁 70-80。
18. 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。
19. 楊秋敏(2016)。〈AI 人工智慧與老人照護〉,《開南法學》,8 特刊,頁 219-242。
20. 廖淑君(2011)。〈智慧聯網之發展與個人資訊隱私保護課題:以歐盟之因應為例(Internet of Things vs. information privacy protection: take European Commission's actions as an example)〉,《科技法律透析》,23 卷 11 期,頁 18-42。
21. 潘俊良(2017)。〈美國白宮公布「為人工智慧的未來準備」報告〉,《科技法律透析》,29 卷 1 期,頁 5-7。
22. 潘俊良(2017)。〈簡析德國自動駕駛與車聯網發展策略〉,《科技法律透析》,29 卷 4 期,頁 25-33。
23. 潘俊良(2018)。〈自駕車之發展與挑戰－以德國法制為借鑑(Development and Challenges of Autonomous Driving－Taking the German legal Issues as a reference)〉,《科技法律透析》,30 卷 12 期,頁 48-72。
24. 蔡碩庭(2018)。〈網路服務提供者侵害著作權之民事責任(Civil Liability for Copyright Infringement of Internet Service Providers)〉,《智慧財產評論》,15 卷 1 期,頁 163-216。
25. 蕭仁豪(2018)。〈日本人工智慧(AI)發展與著作權法制互動課題之探討〉,《科技法律透析》,30:1 期,頁 46-72。
26. 賴俊傑(2009)。〈迎接一個人與機器人共存的社會－日本次世代機器人安全性確保準則之淺介〉,《科技法律透析》,21 卷 6 期,頁 2-7。
27. 闕光威(2008)。〈E 世代知識管理管理平臺中隱私權與智慧財產權法的爭議 (A New Mode of Knowledge Transfer and the Relevant Intellectual Property and Privacy Issues in Taiwan)〉,《智慧財產評論》,6 卷 2 期,頁 79-92。
28. 羅文妙(2018)。〈淺談機器人及其專業佈局〉,《理律法律雜誌雙月刊》,107 年 1 月號,頁 9-10。

專書論文

1. 劉靜怡(2018)。〈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〉,收於:劉靜怡(主編),《人工智慧相關法律議題芻議》,頁 3-45。台北:元照。
2. 顏厥安(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(主編),《人工智慧相關法律議題芻議》,頁 50-85。台北:元照。
3. 吳從周(2018)。〈初探 AI 民事責任—聚焦反思臺灣之實務見解〉,收於:

劉靜怡(主編)，《人工智慧相關法律議題芻議》，頁 89-116。台北:元照。
4. 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(主編)，《人工智慧相關法律議題芻議》，頁 120-148。台北:元照。
5. 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(主編)，《人工智慧相關法律議題芻議》，頁 153-175。台北:元照。
6. 沈宗倫(2018)。〈人工智慧科技與智慧財產權法制的交會與調和—以著作權法與專利法之權力歸屬為中心〉，收於：劉靜怡(主編)，《人工智慧相關法律議題芻議》，頁 181-214。台北:元照。
7. 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(主編)，《人工智慧相關法律議題芻議》，頁 219-241。台北:元照。

(二)英文部分

政府報告

White House Report (2016). Artificial Intelligence, Automation, and the Economy, *available at* https://obamawhitehouse.archives.gov/sites/whitehouse.gov/files/documents/Artificial-Intelligence-Automation-Economy.PDF (Dec. 2016)

研究報告

1. Crawford, Kate et al. (2016). The AI Now Report: The Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term, *available at* https://artificialintelligencenow.com/media/documents/AINowSummaryReport_3_RpmwKHu.pdf (A summary of the AI Now public symposium, hosted by the White House and New York University's Information Law Institute, July 7th, 2016)
2. The One Hundred Year Study on Artificial Intelligence, Report of the 2015 Study Plan (2016). Artificial Intelligence and Life in 2030, *available at*https://ai100.stanford.edu/sites/default/files/ai_100_report_0831fnl.pdf (Sept. 2016)
3. Broad Agency Announcement: Explainable Artificial Intelligence (XAI) DARPA-BAA-16-53 (2016). http://www.darpa.mil/attachments/DARPA-BAA-16-53.pdf (Aug. 2016)

專書

1. Calo, Ryan, A. Michael Froomkin & Ian Kerr eds. (2016). Robot Law. Cheltenham, UK/ Northampton, MA: Edward Elgar Publishing.
2. Ezrachi, Ariel & Maurice E. Stucke (2016). Virtual Competition: The Promise and Peril of the Algorithm-Driven Economy. Cambridge, MA: Harvard University Press.
3. O'Neil, Cathy (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. New York: Crown.
4. Stucke, Maurice E. & Allen P. Grunes (2016). Big Data And Competition Policy. New York: Oxford University Press.
5. Susskind, Richard & Daniel Susskind (2016). The Future of the Professions: How Technology Will Transform the Work of Human Experts. New York: Oxford University Press.
6. Domingos, Pedro (2015). The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World. New York: Basic Books.
7. Hildebrandt, Mireille (2015). Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology. Cheltenham, UK/Northampton, MA: Edward Elgar Publishing.
8. Markoff, John (2015). Machines of Loving Grace: The Quest for Common Ground Between Humans and Robots. New York: HarperCollins Publishers.
9. Pasquale, Frank (2015). The Black Box Society: The Secret Algorithms That Control Money and Information. Cambridge, MA: Harvard University Press.
10. Bostrom, Nick (2014). Superintelligence: Paths, Dangers, Strategies. New York: Oxford University Press.
11. Brynjolfsson, Erik & Andrew McAfee (2014). The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies. New York: W.W. Norton & Company.
12. Quinlan, J. Ross (2014). C4.5: Programs for Machine Learning. Elsevier.
13. Lin, Patrick, George Bekey, & Keith Abney eds. (2014). Robot Ethics: The Ethical and Social Implications of Robotics. Cambridge, MA: The MIT Press.
14. Vapnik, Vladmir N. (2013). The Nature of Statistical Learning Theory. New York, USA: Springer-Verlag.
15. Pagallo, Ugo (2013). The Laws of Robots: Crimes, Contracts, and Torts. New York, USA: Springer-Verlag.
16. Chopra, S. & White, L. F. (2011). A Legal Theory for Autonomous Artificial Agents. Ann Arbor, Michigan: University of Michigan Press.
17. Caruana, R. (1998). Multitask Learning. New York, USA: Springer-Verlag.
18. Sestito, S. & Dillon, T. S. (1994). Automated Knowledge Acquisition, Sydney:

Prentice Hall.
19. Minsky, M., & Papert, S. (1969). Perceptrons. Oxford, England: The MIT Press.

期刊論文與會議論文

1. Cogliance, Cary & David Lehr (2017). Regulating by Robot: Administrative Decision Making in the Machine-Learning Era. *Georgetown Law Journal* (Forthcoming), *available at* http://scholarship.law.upenn.edu/faculty_scholarship/1734/.
2. Ingles, Ignatius Michael (2017). Regulating Religious Robots: Free Exercise and RFRA in the Time of Superintelligent Artificial Intelligence. *Georgetown Law Journal* 105:507-530.
3. Alarie, Benjamin et al. (2016). Law in the Future. *University of Toronto Law Journal* 66: 423-428.
4. Alarie, Benjamin (2016). The Path of the Law: Towards Legal Singularity. *University of Toronto Law Journal* 66: 443-455.
5. Brown, Shannon (2016). Peeking Inside the Black Box: A Preliminary Survey of Technology Assisted Review (TAR) and Predictive Coding Algorithms for eDiscovery. *Suffolk Journal of Trial & Appellate Advocacy* 21:221-287.
6. Calo, Ryan (2016). Robots as Legal Metaphors. *Harvard Journal of Law & Technology* 29: 209-237.
7. Casey, Anthony J. & Anthony Niblett (2016). Self-Driving Laws. *University of Toronto Law Journal* 66: 429-442.
8. Castel, Matthew E. & J.-G Castel (2016). *The Impact of Artificial Intelligence on Canadian* 46:34-59.
9. Etzioni, Amitai & Oren Etzioni (2016). Keeping AI Legal. *Vanderbilt Journal of Entertainment and Technology Law* 19:133-146.
10. Fraser, Erica (2016) Computers as Inventors - Legal and Policy Implications of Artificial Intelligence on Patent Law. *SCRIPTed: A Journal of Law, Technology and Society* 13:305-333.
11. Joh, Elizabeth E. (2016). Policing Police Robots. *UCLA Law Review Discourse* 64:516-543.
12. Khan, Fazal (2016). The "Uberization" of Healthcare: The Forthcoming Legal Storm over Mobile Health Technology's Impact on the Medical Profession. *Health Matrix* 26:123-172.
13. Massaro, Toni M. & Helen Norton (2016). Siri-Ously? Free Speech Rights and Artificial Intelligence. *Northwestern University Law Review* 110:1169-1194.
14. Obermeyer, Ziad & Ezekiel J. Emanuel (2016). Predicting the Future — Big

Data, Machine Learning, and Clinical Medicine. *The New England Journal of Medicine*, (September 29, 2016). *Available at* http://www.nejm.org/doi/full/10.1056/NEJMp1606181.

15. Paliwala, Adbul (2016). Rediscovering Artificial Intelligence and Law: An Inadequate Jurisprudence. *International Review of Law, Computers & Technology* 30:107-114.

16. Rothenberg, David Marc (2016). Can Siri 10.0 Buy Your Home? The Legal and Policy Based Implications of Artificial Intelligent Robots Owning Real Property. *Washington Journal of Law, Technology & Arts* 11:439-460.

17. Schafer, Burkhard (2016). Editorial: The Future of IP Law in an Age of Artificial Intelligence. *SCRIPTed: A Journal of Law, Technology & Society* 13:283-288.

18. Scherer, Matthew U (2016). Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies. *Harvard Journal of Law & Technology* 29:353-400.

19. Simshaw, Drew, Nicholas Terry, Dr. Kris Hauser, & M.L. Cummings (2016). Regulating Healthcare Robots: Maximizing Opportunities While Minimizing Risks. *Richmond Journal of Law & Technology 22:3.* http://jolt.richmond.edu/v22i2/article3.pdf.

20. Surden, Harry & Mary-Anne Williams (2016). Technological Opacity, Predictability, and Self-Driving Cars. *Cardozo Law Review* 38: 121-181.

21. Talley, Nancy B. (2016). Imagining the Use of Intelligent Agents and Artificial Intelligence in Academic Law Libraries. *Law Library Journal* 108:383-402.

22. Villasenor, John (2016). Technology and the Role of Intent in Constitutionally Protected Expression. *Harvard Journal of Law & Public Policy* 39:631-674.

23. Yoon, Albert H (2016). The Post-Modern Lawyer: Technology and the Democratization of Legal Representation. *University of Toronto Law Journal* 66:456-471.

24. Zimmerman, Larry N. (2016). Artificial Intelligence in the Judiciary. *Journal of the Kansas Bar Association* 85:20-23.

25. Balkin, Jack M. (2015). The Path of Robotics Law. *California Law Review Circuit* 6:45-60.

26. Hattenbach, Ben & Joshua Glucoft (2015). Patents in An Era of Infinite Monkeys and Artificial Intelligence. *Stanford Technology Law Review* 19:32-51.

27. Calo, Ryan (2015). Robotics and the Lessons of Cyberlaw. *California Law Review* 103:513-564.

28. Jones, Meg Leta. (2015). The Ironies of Automation Law: Tying Policy Knots

with Fair Automation Practices Principles. *Vanderbilt Journal of Entertainment and Technology Law* 18:77-134.

29. Ferguson, Andrew Guthrie (2015). Big Data and Predictive Reasonable Suspicion. *University of Pennsylvania Law Review* 163: 327-410.

30. Myers, Laura, Allen Parrish, & Alexis Williams (2015). Big Data and the Fourth Amendment: Reducing Overreliance on the Objectivity of Predictive Policing. *Federal Courts Law Review* 8:231-.

31. Oancea, Cristian-Vlad (2015). Artificial Intelligence Role in Cybersecurity Infrastructures. *International Journal of Information Security & Cybercrime* 4:59-62.

32. Pasquale, Frank & Glyn Cashwell (2015). Four Futures of Legal Automation. *UCLA Law Review Discourse* 63:26-48.

33. Proia, Andrew, Drew Simshaw & Kris Hauser (2015). Consumer Cloud Robotics and the Fair Information Practice Principles: Recognizing the Challenges and Opportunities Ahead. *Minnesota Journal of Law, Science & Technology* 16:145-214.

34. Raymond, Anjanette H. & Scott J. Shackelford (2015). Jury Glasses: Wearable Technology And Its Role in Crowdsourcing Justice. *Cardozo Journal of Conflict Resolution* 17:115-154.

35. Scopino, Gregory (2015). Preparing Financial Regulation for the Second Machine Age: The Need for Oversight of Digital Intermediaries in the Futures Markets. *Columbia Business Law Review* 2015:439-519.

36. Sheppard, Brian (2015). Incomplete Innovation and the Premature Disruption of Legal Services. *Michigan State Law Review* 2015:1797-1910.

37. Thompson, Darin (2015). Creating New Pathways to Justice Using Simple Artificial Intelligence and Online Dispute Resolution. *International Journal of Online Dispute Resolution* 2:4-53.

38. Van Arsdale, Suzanne (2015). User Protections in Online Dispute Resolution. *Harvard Negotiation Law Review* 21:107-142.

39. Citron, Danielle Keats & Frank Pasquale (2014). The Scored Society: Due Process for Automated Predictions. *Washington Law Review* 89:1-33.

40. Eszteri, Daniel (2014). Responsibility for Damages Caused by Artificial Intelligence. *PhD tanulmanyok* 13:123-158.

41. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Aaron, C. & Bengio, Y. (2014). Generative Adversarial Nets. *Advances in Neural Information Processing Systems* 2672-2680.

42. Joh, Elizabeth E. (2014). Policing by Numbers: Big Data and the Fourth Amendment. *Washington Law Review* 89:35-68.

43. Katz, Pamela S. (2014). Expert Robot: Using Artificial Intelligence to Assist Judges in Admitting Scientific Expert Testimony. *Albany Law Journal of Science & Technology 24:1-46.*

44. McGinnis, John O. & Russell G. Pearce (2014). The Great Disruption: How Machine Intelligence Will Transform the Role of Lawyers in the Delivery of Legal Services. *Fordham Law Review* 82: 3041-3066.

45. Stewart, Chip (2014). Do Androids Dream of Electric Free Speech? Visions of the Future of Copyright, Privacy and the First Amendment in Science Fiction. *Communication Law and Policy* 19:433-463.

46. Surden, Harry (2014). Machine Learning and Law. *Washington Law Review* 89:87-115.

47. Vladeck, David C. (2014). Machines without Principals: Liability Rules and Artificial Intelligence. *Washington Law Review* 89:117-150.

48. Calo, M. Ryan (2011). Open Robotics. *University of Maryland Law Review* 70: 571-613.

49. Huang, L., Joseph, A. D., Nelson, B., Rubinstein, B. I., & Tygar, J. D. (2011). Adversarial Machine Learning. *Proceedings of the 4th ACM Workshop on Security and Artificial Intelligence* 43-58.

50. Kobayashi, Bruce H. & Larry E. Ribstein (2011). Law's Information Revolution. *Arizona Law Review* 53: 1169-1220.

51. Pan, S. J., & Yang, Q. (2010). A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering.* 22(10) : 1345-1359.

52. Aksoy, M. S. (2008). A Review of Rules Family of Algorithms. *Mathematical and Computational Application* 13(1) : 51-60.

53. Citron, Danielle K. (2008). Technological Due Process. *Washington University Law Review* 85: 1249-1318.

54. Allen, Colin, Windell Wallach & Iva Smit (2006). Why Machine Ethics? *IEEE Intelligent Systems* 21: 12-17.

55. Li, F., Rob, F., & Perona, P. (2006). One-Shot Learning of Object Categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 28(4) : 594-611.

56. Roweis, S. T., & Saul, L. K. (2000). Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science* 290(5500), 2323-2326.

57. Sutton, R. S., & Barto, A. G. (1998). Reinforcement Learning: An Introduction (Vol. 1, No. 1). Cambridge, MA: MIT Press.

58. Wolpert, D. H. (1996). The Lack of a Priori Distinctions Between Learning Algorithms. *Neural Computation* 8(7): 1341-1390.

59. Fu, L. (1994). Rule Generation from Neural Networks. *IEEE Transactions on*

Systems, Man, and Cybernetics 24(8) : 1114-1124.

60. Towell, G. G., & Shavlik, J. W. (1994). *Knowledge-Based Artificial Neural Networks.* Artificial intelligence 70(1-2) : 119-165.

61. Fu, L. (1991). Rule Learning by Searching on Adapted Nets. *Proceedings of the Ninth National Conference on Artificial Intelligence* 91: 590-595.

62. Harnad, S. (1990). The Symbol Grounding Problem. *Physica D: Nonlinear Phenomena.* 42(1-3): 335-346.

63. McCarthy, J., & Hayes, P. J. (1969). Some Philosophical Problems from the Standpoint of Artificial Intelligence. *Readings in artificial intelligence,* 431-450.

(三)德文部分

專書

1. Eric Hilgendorf/Jochen Feldle.(2018). Digitization and the Law.
2. Lennart S. Lutz.(2017). Automatisiertes Fahren, Dashcams und die Speicherung beweisrelevanter Daten.
3. Florian Münch.(2017). Autonome Systeme im Krankenhaus.
4. Eric Hilgendorf.(2017). Autonome Systeme und neue Mobilität.
5. Raphael Klesen.(2017). Die Entscheidung von Maschinen über Menschenleben
6. Lisa Blechschmitt.(2017). Die straf- und zivilrechtliche Haftung des Arztes beim Einsatz roboterassistierter Chirurgie.
7. Eric Hilgendorf/Uwe Seidel.(2017). Robotics, Autonomics, and the Law.
8. Melinda Florina Lohmann.(2016). Automatisierte Fahrzeuge im Lichte des Schweizer Zulassungs- und Haftungsrechts.
9. Sabine Gless/Kurt Seelmann.(2016). Intelligente Agenten und das Recht.
10. Susanne Beck/Bernd-Dieter Meier/Carsten Momsen.(2015). Cybercrime und Cyberinvestigations.
11. Eric Hilgendorf/Sven Hötitzsch.(2015). Das Recht vor den Herausforderungen der modernen Technik.
12. Eric Hilgendorf/Sven Hötitzsch/Lennart S. Lutz.(2015). Rechtliche Aspekte automatisierter Fahrzeuge.
13. Eric Hilgendorf.(2014). Robotik im Kontext von Recht und Moral.
14. Eric Hilgendorf/Jan-Philipp Günther.(2013). Robotik und Gesetzgebung.
15. Susanne Beck.(2012). Jenseits von Mensch und Maschine.