公民．價值 @AI 論壇

時　　間：108 年 10 月 29 日（二）13:00-17:30
地　　點：國立臺灣大學圖書館國際會議廳
記　　錄：李仲軒（科技部人文社會科學研究中心博士後研究員）

本次論壇由科技部人文司鄭毓瑜司長開場引言，開宗明義指出人文社會研究不是在發明 AI 科學技術，但是 AI 科技的運用一定影響每一個人的生活習慣、行為思想，也影響社會、強化社會控制，因此保持警覺，關注 AI 科技的發展，與科技展開交涉或對話，也是人文社會研究的重要項目。

然而在此之前，人文社會研究也需要反思，科技發展可以或應該被限制嗎？人類社會的普遍價值還有什麼？人類世到後人類，人要如何重新定義？人文科學的基礎價值在哪裡？這些議題都需要大家一齊來集思廣益。今天的論壇亦將介紹科技部人文司所推動的兩項專案計畫：「人工智慧的創新與規範：科學技術與人文社會科學的交互作用」及「臺灣人文社會的價值基礎：多元性與價值衝突的反思與研究」，希望建立討論平臺，展開對話。

場次一：AI的葫蘆裡到底賣什麼藥？

主持人：蔡政宏（中央研究院歐美研究所研究員）
主講人：李建良（中央研究院法律學研究所研究員）
　　林文源（國立清華大學通識教育中心教授）
　　邱文聰（中央研究院法律學研究所副研究員）

一、李建良教授談〈人工智慧與人性尊嚴〉

李建良教授以「人工智慧與人性尊嚴」為題，從「人工智慧無所不能？人性尊嚴不容侵犯！」開始，以一個問號和一個驚嘆號揭開認識論與規範論的思考方向與交互作用。人工智慧可能由弱 AI 發展成強 AI，這引起人們的期待、恐懼

本文由李仲軒博士記錄整理，經各場次主講人審訂。

甚至焦慮。相應於此，對「人性」或「人性尊嚴」的信念堅持是不是也要能相對的增強？換言之，不只 AI 科技需要創新與規範，法律本身同樣也要創新與規範。法律在與新興科技互動演進的過程中，需要受到更上位的人性尊嚴，作為價值的規範與指引。

近期科技部公布的人工智慧科研發展指引，基本上就是這樣的一種嘗試。暫且不問人性尊嚴如何定義，該指引明文表彰人性尊嚴的價值，整份科研發展指引可以被視為是體現人性尊嚴價值的思維框架，足以建立思考架構的起點。該份指引作為「目標設定」，確立了人性尊嚴及相關的價值，但還需要透過「行為準則」、「落實機制」的部分來具體落實人性尊嚴。

回到最初的選題，認識論的重點在探究 AI 能做到什麼？對人類有何益處？有何傷害？這必須先分別「公部門」與「私領域」兩種面向進行分析，因為這兩者的侵害類型或程度可能有所不同。在這之後，才可能進行規範論層次的探索，包含法律在內的規範論思考。首先必須進行價值的選擇，幾種潛在的價值選項，也可能會相互衝突矛盾。例如，增強競爭力和人的自主性，可能都是值得追求的價值，也都可以衍生出不同的規範內容。但是，這兩種價值選擇，彼此間可能存在衝突，這方面的釐清，需要更多的討論，期待和另一個 AI 臺灣價值團隊就此可以有更多交流。

關於價值設定，一個初步的想法是「安全與自由」、「理性與尊嚴」，四種價值、兩組分類，構成多種組合。其彼此之間，可能存在衝突。但是，在針對 AI 的發展時，此四者卻也可以是正向回饋的關係。人類無法阻擋 AI 的發展，但 AI 的發展如能確保其安全性，人的自由、自主性就不至於喪失，人的尊嚴也能

圖一：公民．價值 @AI 論壇

受到維護、免於侵犯。反過來說，要確保 AI 發展的安全性，首先就必須重視人的尊嚴、自由與自主性。

在另一個面向上，當人們越能運用理性，就越能建構一個理性的思考框架，以發展值得信賴、可靠的 AI 以實現安全、自由與尊嚴。但是人們是不是真的有理性？人如果不理性，如何達成理性？或許對於 AI 問題的人文社會研究，最終會回到人自身的問題。「人工智慧讓人越來越像人，還是越來越不是人？」重點或許不只是要創造值得信賴跟可靠的人工智慧，更重要的是，如何實現人的尊嚴，讓人類真正能夠運用理性，讓人類自己更加可靠、值得信賴。

二、林文源教授談〈由技術物的政治性到 AI 的公民性格〉

清大通識教育中心林文源教授，以 AI for the 99 為題，表現由「技術物的政治性」到「公共化 AI」的企圖心。林教授首先以社會學及科技與社會研究（STS）觀點為例，闡述 AI 與人文社會（HSS）的四種相互影響關係：

HSS in AI（實然面）：AI 造成的人社議題！
HSS of AI（分析面）：人社研究如何剖析 AI ？
HSS by AI（方法面）：AI 如何開拓人社研究？
HSS for AI（應然面）：人社研究如何拓展 AI 願景？

AI 作為技術物（artifact）的政治性，是指依據 Langdon Winner 的觀點，AI 作為包含大數據、演算法、機器學習、詮釋與應用的技術集合，會造就人類社群連結中的權力與權威安排，並改變其中發生的活動。

在政治、治理場域，不論是集權政體或民主政體，AI 都可以強化治理權力，中國社會信用評分及 Amazon Rekognition 中的 Face surveillance，皆為著例。在經濟、市場場域，AI 影響資本生產與分配，以數據資本主義的型態，促進經濟發展與轉型。AI 不只驅動行銷，AI 還已經介入影響整體企業，涵蓋藍領、白領到決策階層，由招募到評量、生產與再生產的活動。在社會場域，因為也逐漸依賴社群媒體與網路傳播，形塑各種群體的社會認知，自然也存在 AI 透過類似 Deepfakes 的方式，侵蝕公民社會、社會信任與民主根基。這些都是由 HSS in AI 的思路所做的反省與批判，由人文社會視野揭露演算法黑箱、批判 AI 的政治性與毀滅性。

但除了批判演算法與資料偏誤、數據代表性、文化差異，或歸罪於政府官僚、資本家、工程師，與數據製造者外，林教授強調，由 HSS 視野連結 AI 與

社群，提供重新想像 AI 與公民社會的可能。這種公共化的 AI，是以 HSS 的公民社會理念與價值，脈絡化數據、演算法、學習、詮釋與應用，提供不同社會（群）、價值進行脈絡的連結與創造新的機會。透過偵測、中介、循環等作用，最終造成轉變。林教授並以世界與臺灣的政府機構、大學研究團隊、非政府組織與商業應用的數種案例，扼要點出幾種介入與轉變的方式。其中所謂轉變包括，將數據連結脈絡，將技術連結社群，將 AI 連結政策分析，最後將理念連結技術社群。這也就是由反省 AI 實然面（HSS in AI）開始，帶入人社分析與想像（HSS of AI），連結 AI 技術面（HSS by AI），以打造契合人社價值的 AI（HSS for AI）。

是以，從社會學與 STS 觀點，AI 不只服務治理與資本（1%），侵蝕公民社會（99%）。AI 也有機會強化社會資本，問題在於：AI 需要人社想像力，以重新想像、連結這些可能性。以 HSS 重新想像公共化的、for the 99%的 AI，重建社會連帶與信任的社會資本。這似乎可以和李建良教授演講最後的提醒相呼應。

三、邱文聰教授談〈誰的人工智慧？模仿與學習的價值預設〉

邱文聰教授以「誰的人工智慧：模仿與學習的價值預設」為題，深入檢視 AI 的技術層面，以釐清這裡所討論、研究的 AI 究何所指。由技術進化的角度分類 AI，可以略分為三個階段。目前所處的第二波 AI 是以資料分析、機器學習、演算法為核心，其固然還不是第三波全能的後人類 AI，但已經強於前一階段 AI，僅是透過機器語言，讓機器學習、操作人類社會活動所已知的科學知識結晶的規則或定律定理，以發揮作用。以下的討論，聚焦在第二波 AI 所引發的問題。至於第三波 AI 所可能出現的更強的全能型超級智能，還不是現階段最迫切的問題。

第二波 AI 不依靠由人類，尤其是專家系統生產的知識作為運行基礎，更不依賴有限的規則、定律，去掌握複雜的真實世界。其本於資料科學與機器學習的技術，提供給機器關於人類活動所產生、遺留的資料越大量，機器即可自行生產越多關於人類的知識。由於此種知識生產具有資料驅動的本質，資料是 AI 掌握人類知識的關鍵，其自然受資料所在的限制，而呈現區塊、任務領域孤立發展的現象。

以 AI 知識生產的型態、方式，或資料的來源，可略分為非個資依賴型及個資依賴型 AI。後者是收集與個人相關的資料作為機器學習的訓練素材，易引起各種法律與倫理上的疑慮。解決這些問題的基本法則是，自主權保障程度應該

與利用之公益性與必要性成反比。唯有重大公益與必要性，可以正當化對自主權的重大限制。

AI 的各種令人意想不到的運用，其實也可以略分為兩類：第一類是模仿型 AI，旨在模仿人類行為，並以此為標竿，著眼於以機器自動化來取代人工、增加效率，但其知識能力並未超越人類，所以人可以量度、評估、控制 AI。第二類的 AI 運用，旨在發現隱藏的模式或現象間未知的關聯性，希望藉由找到關聯性變因，從而影響所欲控制的結果或達成之目標。關聯性的知識或關聯性推論當然還是與因果關係知識不同，不能完全填補知識缺口，但是卻可以幫助人們作成決定、決策，強化控制或治理。所以雖然這類的 AI 運用，看似僅是中性的知識探索，但是其背後還是有明確的目的方向甚至價值的選擇與設定。

當面對 AI 這種可以增加效率、增強控制的工具，應該透過科學哲學家 Philip Kitcher 所指出的三個層次的思辨，來決定應如何運用 AI：首先是決定價值體系順位，要優先達成何種目標，並擇定具體的議題。其次，須本於認知體系，選擇作出觀察或提出問題的角度。最後，則是在證據體系探究是否應該容許運用關聯性推論，而非明確因果證據，來做出行動的決定：有時候，要絕對避免第一型的偽陽性錯誤（false positives），有時候卻應該優先避免第二型的偽陰性錯誤（false negatives），這都必須要在具體的類型、脈絡下，才能做成判斷。

所有這些價值的決定，絕對不是技術專家所獨擅的場域，而是要靠社會共同思辨。這些構成 AI 的公民、價值議題，也就是今天論壇的主題。

場次二 ：AI對人文社會價值的衝擊

主持人：鄧育仁（中央研究院歐美研究所研究員）
主講人：黃冠閔（中央研究院中國文哲研究所研究員）
　　賴曉黎（國立臺灣大學社會學系副教授）
　　廖朝陽（國立臺灣大學外國語文學系教授）

一、黃冠閔教授談〈「誰的」人工智慧？價值衝擊與重估〉

下半場由黃冠閔教授由價值衝擊與重估的角度，提問人工智慧究竟是屬於「誰的」？其所帶領的科技部人文司「臺灣人文社會的價值基礎：多元性與價值衝突的反思與研究」團隊，對於 AI 提供人文價值思考為基礎的戰略思考。自二戰後，臺灣都以尋求生存方案為優先，在知識創新與安身立命上以遵循既定模

式為最能確保生存的方法。但隨著全球化、差異化、多元化的趨勢衝擊，臺灣必須找到新的生存模式，甚至是新的價值衡量基準。如何對人類存在的價值、生活品質提出貢獻，是人文價值可提供的視野，以人類共同生活的福祉為前提，回應人類生活的危機與希望。

這種價值反思的切入點具有多元性，包括知識分工、利益目標分化、認同與價值選擇的多樣化，並且包含逐步擴大，由個人到群體、物種、甚至星球的思考結構，以回應歷史性與當代性的提問。導入價值思考，是因為舉凡 AI 治理原則及規範制定，都必須存在價值基礎的共識。

AI 在演算的機器架構上，進行速度、規模、標準、型態的競爭，不僅是國家競爭的戰略指標，也成為新的資本累積型態。不僅產生新的資本型態（資本與資料的結合），也形成新的勞動關係（矽基勞動力與電力結合）。

要鞏固 AI 優勢所築起的技術、資本、經濟規模、資料量的門檻越來越高，投入、參與所需的資源就越來越多。AI 被握有資本、權力掌控數位科技知識的少數人掌握，資本與權力將更為集中，實現數位科技民主化的同時，也產生更多被排除的「數位賤民」與不平等，形成新的宰制關係，政治、法律、倫理、社會關係都將重新形塑。所以最根本的價值問題是，AI 究竟應該是「誰的」？

二、賴曉黎教授談〈資通科技的兩面性——從工具人與遊戲人談起〉

賴曉黎教授首先引用 Georg Simmel「客觀文化凌駕於主觀文化之上」的觀點，認為人永遠跟不上客觀文化，是永恆的落伍者，不可能預測人類的未來趨勢。與其嘗試掌控科技，不如處理人們實實在在的焦慮、不安的感受。這需要我們進一步反省，人活著的目標什麼？人之所以為人的本質是什麼？現代社會應如何看待新的技術物或建立新的關係？

一是以勞動的角度來看，就是對世界由目的性的角度加以改造。人類發明機器、新科技的意義，是為了增加勞動的效用、效率，擺脫必然性的控制，不要被自然奴役。而這真的就是人活著的唯一目標或人的本質嗎？是我們需要的社會或未來嗎？

另一種可能性是用互動、或是遊戲的角度來設想。遊戲根本上就是人先主動參與、設立非必要的障礙，再樂此不疲地加以克服的活動。作為社會中的人，並不是非要天天勞動、只講效率。人的本質或價值，應該有超越功用性、

利益、效率等工具思維的部分，例如社交、溝通、對話等非功利性，非目的性、不需要計算利益的活動。

正如狄更斯講的「這個時代是最好的時代，也是最壞的時代」，端看人們如何應對、選擇。從勞動到互動，從主體到相互主體，從工具人到遊戲人，正是一個未來世代必然要面對的價值選擇的問題。而對社會研究者而言，沒有所謂客觀存在的價值。價值永遠是一個動詞，是通過互動、通過討論，甚至經過競爭，鬥爭產生評價的過程。

我們可以選擇繼續做勞動工具人，不斷提高生產力、效率以滿足需求、增加控制。但是這在當今過度生產的消費社會，有何意義？當人越想要剝削、壓榨對方，結果反而會讓人類自身陷入更大的恐懼與苦難的夢魘，擔心機器人會越做越好，許多職業即將消失，有了自主意識後如果造反，人類的未來在哪裡？當人類把科技當成僕人、工具來利用，以控制的方式增加效率，最終就會把科技變成人類的敵人。

但其實這一切毋須如此。只要我們對人或新的世代有更多的認識與信心，不要用現有的權力，去強制他們用我們既有的思考方式去面對未來。我們的思考方式必須讓他消失在歷史的灰燼中，價值是人集體創造的，如尼采所說「永恆回歸」與「生成的無辜」：儘管每次選擇都跟過去一樣，但每次都是從頭開始新的選擇。

如同以禮物文化為基礎原型的網路的出現，就不是基於功利性、利益性、目的性，只是單純的為了達到溝通、對話的互助、共享。新的世代希望用互動、遊戲的角度，重新面對不一樣的世界。我們不要妨礙人類繼續創造新的價值，來取代現在的價值。面對新世界，一定需要新的合作方式，新的面對世界的方式。

三、廖朝陽教授談〈後人類思維與人文技術的未來性〉

最後由廖朝陽教授來談，後人類思維與人文、技術的未來性。人工智慧的問題，在真正的通用人工智慧出現前，本質上還是人與技術之間使用關係的問題。

第一個問題是技術等於能力的外置。依照 Bernard Stiegler 的理解，所有技術本質上自始就是人的能力的外置。人工智慧英文是 intelligence，翻譯成「智慧」已經不是傳統的意思而是指向實用的，解決問題的能力。如此一來，所有技術都牽涉到智慧。既然技術都是人工創造的，那是不是所有人的技術發明都是

人工智慧？進一步來想，人體的內部器官，例如心臟，是不是也是一種能力的外置，從而也是技術，也就是自有一套演算法的「天然」智慧器官？從這個角度看，人工智慧的問題必須從更大的物種特性來考慮，特別是在技術工業化以後，技術發展走向毒化，產生人的能力被剝奪、獨立判斷能力萎縮等現象。人工智慧的問題只是其中的一部分。

當人與技術物之間的區隔趨向模糊，我們會發現森政弘所提出的恐怖谷假說可以有另一種意思。恐怖谷假說原本是說機器人模擬真人，模仿得非常像時，一點點不像都會引起親和感驟降。但是森的圖解除了實用的技術物之外也將各種人造物（文樂人偶、佛像雕刻等等）擺在同一軸線上，指向人文也是一種技術。原本森的意思是形似往上提升會導向神似，類似弱 AI 轉向強 AI。但我們也可以反過來，把恐怖谷的另一側（更高的親和感）看成人文知識底層的技術性，其中的神似仍然是形似的延伸。

這就是由後人類觀點來重新理解智慧的問題。能力是在個體的內部還是外部，通常可能會被視為指向意識與技術的區分，這是身心分離的觀點，也就是被看成人類最重要能力的理性才是個體的核心，身體則是某種外部機器。後人類就是要去改變既有的思維習慣，把人類物種能力的核心（神似）看成外置能力（弱 AI 的形似）的延伸。這就表示物際關係也可以向演算法之外延伸，指向恐怖谷對面的高峰。森政弘在這裡看到機器人的佛性。我們可以在這裡看到透過技術本身來恢復長循環、多面向能力，化解技術毒化的可能。

圖二：綜合討論（左起為邱文聰副研究員、林文源教授、李建良研究員、黃冠閔研究員、賴曉黎副教授、廖朝陽教授）

綜合討論

主持人：李建良（中央研究院法律學研究所研究員）
　　黃冠閔（中央研究院中國文哲研究所研究員）

最後綜合討論的環節中，有資工系教授談到，進行人文社會科學與 AI 的研究，須注意 AI 是專有名詞，有其固定內容。大數據，或是相關性，屬於統計領域，和 AI 尚有所不同。但重點是，近來日漸對 AI 感到恐懼擔憂，類神經網路運算所得出的結果，完全是人類完全無法理解的。邱文聰教授的回答是，對於深度學習的神祕性，技術上的回應方式是創造可解釋的（Explainable）的 AI，而法律途徑可能要立基於「要求提出說明」的權利。總體來說，從來賓專注的眼神，和熱切的提問不難發現，這次的工作坊充分切合社會與學術研究的需要。兩個計畫團隊分別由科技所可能引發的倫理與法律上議題回溯人文價值，以及從臺灣在地多元的人文價值出發，探索該如何回應 AI 科技的新發展。這代表兩個相輔相成的人文社會研究面向，並於最後一起面對不同學術背景的來賓提問，形成很好的論辯與互動結構。讓所有與會者體認到，對於 AI 的研究，不只關於其可能帶來的效能運用的期待、對其可能帶來風險的恐懼、最終還有對全體人類終極的拷問：究竟人是什麼？人類之異於 AI 者幾兮？從而這次的工作坊，不僅能使我們更了解 AI，其實也是使我們更了解自己。