人文社會科學與 AI：跨領域合作研習營[#]

時　　間：110 年 11 月 19 日（五）13:00-17:00
地　　點：國立清華大學旺宏館二樓 R245 周懷樸講堂
主 講 人：余貞誼（高學醫學大學性別研究所助理教授）
　　韓幸紋（淡江大學會計學系教授）
　　曾正男（國立政治大學應用數學系副教授）
　　李昆樺（國立清華大學教育心理與諮商學系助理教授）
　　區國良（國立清華大學學習科學與科技研究所副教授）
　　王道維（國立清華大學物理學系教授）
主 持 人：王一奇（國立中正大學哲學系暨研究所教授、科技部人文司哲學學門召集人）
　　杜文苓（國立政治大學創新國際學院院長、科技部人文司科技、社會與傳播學門召集人）
　　陶振超（國立陽明交通大學傳播與科技學系教授）
　　連賢明（國立政治大學財政學系教授）
記　　錄：邱冠瑜（國立清華大學社會學研究所研究生）

一、活動介紹

人工智能（Artificial Intelligence, AI）是二十一世紀備受矚目的重要科技。面對 AI 已經啟動的各項改變，跨領域對話與合作勢在必行。人文社會科學領域要如何參與？如何讓 AI 更為趨近人文社會的理念？人社與 AI 領域的有志者要如何合作呢？本次研習營由科技部人文社會科學研究中心第三次委託國立清華大學人文社會 AI 應用與發展研究中心主辦，分成四個場次：第一場「用數據實現性別正義，數據計畫的倫理分析」，由高雄醫學大學性別研究所余貞誼助理教授主講；第二場「查審流程自動化機器人計畫案跨領域合作經驗分享」，由淡江大學會計學系韓幸紋教授與國立政治大學應用數學系曾正男副教授依序主講；第三場「社群媒體下的求助訊號」，由國立清華大學教育心理與諮商學系李昆樺助

[#] 本文由邱冠瑜研究生記錄整理，經國立清華大學人文社會 AI 應用與發展研究中心林文源主任審訂。

理教授與學習科學與科技研究所區國良副教授依序主講；第四場「人文社會 AI 導論課程的設計理念與應用方式」，由國立清華大學物理學系王道維教授主講並為研習營做總結。

本次研習營較先前舉辦場次更為特殊之處，在於國立清華大學人文社會 AI 應用與研究中心結合本研習營與人文司之支持，邀請中心副主任王道維教授拍攝一系列十集之「人文社會 AI 導論」線上課程。[1] 這系列課程一上線即受到相當關注，顯示有相當廣泛的學科都認為 AI 推動與發展有其意義與重要性，這也成為參加研習營者的事先預習與後續學習，讓本活動有更為豐富的層面與長遠影響。

二、用數據實現性別正義，數據計畫的倫理分析

第一場次由王道維教授開場致詞，提及 AI 對我們的生活帶來許多改變，影響人們的生活甚廣，因此在 AI 的運用上需要有人文方面的思考，加深 AI 與人之間的聯繫和合作。在 AI 的應用上，也需要更多跨領域的研究來加深我們對 AI 在不同面向的了解，寄望此次的研習營能為現場的來賓帶來啟發性的思考。

本場次主講人余貞誼教授的研究專長是性別研究及資訊社會學，為大家帶來數據的倫理初探，以及探討 AI 應用在性別議題上的現況，讓與會者更能看見在 AI 應用上應該要留意的部分。首先談到數據計畫，提出兩個問題讓我們思考：如何理解數據計畫？數據計畫與客觀性以及權力之間的關係又是如何？要討論何謂數據計畫可以從三個觀點出發：（1）本體論，把事物轉成後設資料的形式，使得資料的意義轉變成能夠依據目的來運算處理的邏輯判準。（2）認識論，以演算法來梳理肉眼未能察覺的模式和結構。（3）方法論，以數學普遍性為基礎，訴諸機械客觀性（mechanical objectivity），來過濾人為偏見與詮釋侷限，成就對「客觀」、「嚴謹」知識的渴望。

有關數據的客觀與嚴謹，余教授進一步說明數據計畫與客觀性，究竟演算法有沒有偏見呢？她引述數據科學家 Julien Lauret 的看法：「演算法沒有能動性、沒有意識、沒有自主權、沒有道德感。他們只做設計者要求他們做的事。」可是當人的設計者去設計演算法時，就把人的目標、人的資源、人的選擇一起放進了整個數據計畫中，於此同時，偏見也就滲入了數據計畫裡。當數據計畫事實上具有偏見時，裡頭就蘊含著權力不平等的關係，如 D'Ignazio & Klein 的主張「數據就是權力（data is power）」，指出若我們沒有注意到數據計畫的政治

[1] 詳見科技部人文會科學研究中心「人文社會科學與 AI：跨領域合作研習營」網站：http://www.hss.ntu.edu.tw/discourse_info.aspx?d=127。

性，那麼數據計畫就有可能會幫助既有的優勢階級鞏固權力基礎，而加深對弱勢族群之間的不平等。余教授舉例，在美國有以公共健康服務數據集來評估哪些家庭的小孩容易受虐的數據計畫，但是富有階級通常使用的是私人醫療，因此公共健康服務系統中收到的數據經常大多來自於勞動階級，這就導致了貧窮父母被過度取樣，進而高估貧窮小孩遭遇家暴的風險。這樣的模型混淆了 parenting while poor 與 poor parenting 兩者的意義，並鞏固了既有的階層優勢，促成了惡性循環迴圈。總結對數據計畫的理解，余教授認為數據計畫涉及的是一種科技—文化政治（techno-cultural politics），而我們應該要對數據計畫採取批判性的技術實作立場去探究符碼如何運作，包括它的過程、行動者、功能、目標及其力量和侷限，藉由理解軟體能夠運作、影響的程度，來重新思考它的展望與方向。

理解數據計畫的性質後，從人文社會的角度看待數據計畫又會是什麼樣子？余教授提出 Data for good 的概念，這是 AI4people 科學委員會極力推動的主張。他們認為有 4+1 個主張是使用演算法重要的倫理概念，包括行善、不傷害、自主、正義，以及可解釋性，其中可解釋性又包含可理解與可問責。余教授接續進行數據計畫案例的探討。

第一，我們希望數據計畫是照妖鏡還是過濾器？余教授認為透過數據計畫我們可以看見以往社會中被掩蓋的性別不正義，像是 1969-2017 年間曼布克獎文學得獎作品中對於性別角色的描寫都有特定的偏好（男生醫生、警察，女性護士、老師），此時的數據計畫就是照妖鏡；另外一個案例就是第二屆性別暴力防治駭客松提出的「性別暴力解碼計畫」，其中一個得獎團隊設計「Poly you」，這款以性別友善為出發點的網站套件，可以將網路中帶有性別歧視性別暴力的訊息翻譯或是取代，置換為較中性的詞彙，結合社群的力量傳播，可以在無形中將性別歧視降低，此時的數據計畫就是過濾器。

第二，數據計畫是保護主義還是缺憾敘事？余教授提出數據計畫的保護主義案例。在印度頻繁的性犯罪熱點，警方加裝 AI 監視器，若偵測到女性的表情變化且 AI 判定為性騷擾就會向最近的警局通報，讓員警能夠及時地抵達現場處理。儘管現實上能夠降低犯罪率，但很難可以樂觀的說透過這樣的數據計畫看到權力與壓迫概念產生改變的潛能，因為在這個數據計畫中，女性被設定為潛在受害者，把女性固著為需要保護的對象，因而形成了缺陷敘事，反而可能會加深在男性霸權中被宰制的關係。

余教授總結演算法在使用不當的情況中會造成社會的不正義，所以要使其更易於管理也更公平，就是確保問題形成與方案制定的過程中，盡可能的擴大

參與群體。因此一個好的數據計畫應當具備透明性，讓大家可以看進數據的黑箱，朝向充分知情理解的過渡階段；且數據工作者應具有反身性，數據工作人員需要思考自己所占的位置帶有那些局部性、又承擔了什麼樣的責任。最後，我們需要制定出問責機制，在事情發生後可以找到應該由誰來負責，讓數據計畫的問題可以從消極監督到積極改善；另外是數據計畫將會套用在人的身上，一定要注意數據的脈絡是如何被抽取的，這樣才會盡可能公平的對待每一個群體。

三、查審流程自動化機器人計畫案跨領域合作經驗分享

第二場次的兩位演講者共同合作 AI 應用在稅務審查的計畫，透過兩位講者在實務工作的經驗分享，讓與會者更能了解 AI 應用的過程中會面臨哪些理工與社會的摩擦，又該用什麼樣的心態去面對。

首先由韓幸紋教授分享她與稅捐機關的合作經驗。臺灣社會的氛圍在稅務層面上傾向以「節省」為主要考量，從個人到企業都希望能盡量少繳稅，也因此稅務人員的工作大多在處理繳稅者是否非法逃稅，但是隨著企業經營型態的多元化，跨國經營增多，臺灣的稅務人員業務複雜繁重，所以有了透過大數據及 AI 來協助稅務人員查審的計畫。韓教授分享她在營所稅的查審經驗，其中查審的方法又分為兩種：電腦選案及人工選案。電腦選案的過程會列出幾項特定的條件，之後在資料庫裡面隨機選案；人工選案則是依賴稅務員人工設定條件（比對相關的稅務資料），但是相當耗時且常常徒勞無功，最有效的還是以檢舉案件為主。至於在計畫執行中有關 AI 的應用，韓教授認為 AI 在稅務審查要能成功應用仰賴需求方和工作團隊，需求方須明確提供需求內容，資料的性質要夠豐富，因為稅務人員在蒐集資料時受到相當大的限制，常會遇到資料上橫跨部門而被以資料保護的理由拒絕，造成分析資料時會出現問題；工作團隊內則因為觸及跨領域合作，在合作的初期需要密集溝通，讓彼此在不同領域的經驗可以相互學習，而不被彼此專業領域內的知識所限制。

接續由處理數學模型和 AI 設計的曾正男教授進行分享。曾教授提到當數據計畫碰觸到「人」就會變得複雜，這是理組的工作者必須要注意的部分；而文組的工作者必須要理解 AI 並不是萬能，不能對 AI 和大數據有錯誤的想像，像是 AI 什麼都能做、資料夠大就會有好結果、設計者能夠解釋 AI 分析的細節等。誠如他在與公部門的合作時，都會遇到長官要求 AI 能解決許多複雜的問題，但 AI 的運作邏輯並不是如此，AI 的判準是機率問題，當事件的機率變化太大時，

AI 就不會有好的分析結果。以稅務查審的案件為例，因為資料的缺漏相當複雜，以至於 AI 在學習的時候會不夠完整，還需要人工二次把資料標籤化，而且就算設計出很精準的模型，得到完整的結果，AI 的設計者也不能解釋「為什麼 AI 這麼說」，以上是首要澄清大家對 AI 的迷思。

在實務工作部分，曾教授說明資料的蒐集過程會遇到行政部門彼此的矛盾，當中包含處理不同部門在保護自己的資料時會對需求方產生抗拒心態，並非只要負責的部門申請就能得到想要的資料，因此常在行政程序上耗時許久，他甚至也遇過對方給的資料不正確，或是無法理解需求方的計畫目標，或資料無法準時送到 code 的設計方；又資料在學習 input 與 output 的過程也要一段時間，所以會出現誤判計畫完成期限的差錯。在面對這些問題時，都必須先想像這些資料大概的圖像會是如何，把 code 寫得差不多，等到資料到手再做微幅的修改。又公部門之間常會有利益衝突，我們的成功可能是別人的失敗（曾教授用語），所以除了 AI 技術的問題，在人際關係的協調和溝通也是理組的工作夥伴需要特別留心的部分。

四、社群媒體下的求助訊號

本場次兩位講者共同合作 AI 在心理學的應用計畫，計畫目的是試圖去找出網路上潛在自殺風險的學生，透過臨床心理的專業和 AI 的學習操作，期望能讓跨領域的合作變得更加實用，及早幫助需要特別輔導和關懷的學生。

首先是李昆樺教授的分享。李教授具有臨床心理學的背景，首次嘗試 AI 在學生身心健康和諮商輔導的應用，此次的合作案是在網路社群媒體 Dcard 上找尋自殺個案的高風險族群。團隊先在 Dcard[2] 文章中抓取資料，試圖找尋文章內的自殺風險因子。李教授的工作是負責判別哪些網路文章的字詞或語句屬於「危險的」（亦即會產生自傷或自殺），之後分類危機程度（輕、中、高）、事件程度（人際、學校、家庭、個人）、是否需要藥物或心理治療，接著交給 AI 的專家去建構運算模型。在 Dcard 上的文章有許多並不是有效的內容，導致在標籤資料的過程需要花較多時間，還有在專業領域中的用詞不同導致理解彼此意思的落差，但跨領域的好處就是 AI 的確能增加效率，以往用人工方式看文章效率低落，且容易忽略一些文章的細節，但隨著資料越來越豐富，就能讓這項技術更加成熟，對臨床心理在判別有風險的個案上會有更實質的幫助。

[2] Dcard（狄卡），是臺灣的社群網路服務網站，目前已開放非大學生一般民眾憑證件註冊。（資料來源：「維基百科，自由的百科全書」：https://zh.wikipedia.org/wiki/Dcard（取用時間：2022 年 1 月 7 日））

第二位講者區國良教授則分享過去以文字探勘分析小說的經驗和觀察。為了理解 Dcard 上的用字用語，區教授先設計 AI 去學習 Dcard 用戶較常閱讀的小說，以 LIWC[3] 情緒辭典進行詞性分析，每個詞性具有 60 個詞性維度，接著讓 AI 去分析文章內容的情緒正反面，進而把文章複雜的空間簡化成單維的正反兩面，但這樣做會有盲點，就是單獨使用字詞是不夠的，還必須加入語句分析。在實際進入 Dcard 分析時，團隊從 55,989 篇文章利用情緒字典排序，取出負面分數最高的 2,480 篇文章作為訓練集，再邀請專家為文章內的句子標註 8 個類別，分成 4 個等級：A（有自傷行為的動作）、B（高度危機可能）、C（中度危機可能）、0（低度危機可能），但即便讓 AI 從詞學習到句，還是有句性上的盲點。例如：我不想和「好了」、這樣就「好了」，以上兩句話 AI 就不能準確的判別出情緒的差別。區教授也提及使用演算法上的不同也會對結果造成區別，並舉例用 NB、SVM、DR 分析同一句話就會有差別。而實務上的困難就是特徵數量不足，特徵不一定具有代表性、危機個案程度比例不平均。例如我們平常在路上看到人的背影也許可以從髮型、身材、穿著去猜測性別，但情緒就無法，而且猜測性別也有錯誤的可能，像是留長髮的男生，所以使用 AI 在搜尋文章上，若是機器學習到的特徵不足，也容易出現把有自殺傾向的個案忽略的情況。

區教授總結，現在專家標註的多是負向與自殺憂鬱有關的文字，但兩者之間還是有些微的差異，所以計畫的目的是為了要從大量的資料中快速篩選可能有自傷危機的個案，語意的區分倒不是計畫的目的，而是讓 AI 能夠快速分類語意屬於正或負來到快速分類的目的。現階段這樣二分類的判斷可以有 97%的準確度，可以有效將網路上的負向訊息準確挑出，在未來希望能研究如何使用 AI 挑出正向、負向的文字來預測整篇文章的危機程度，並發展類似二分類模型來更準確找出負向的標籤，協助心理方面在大數據的分析研究。

五、人文社會 AI 導論課程的設計理念與應用方式

在最後的場次，王道維教授為我們帶來 AI 公共化的推廣經驗，透過較為入門的 AI 課程設計，試圖讓 AI 不再限於理工學生專門的技術，其他領域的學生在有相關基礎後也可以一起參與 AI 的使用，進而拓展自身的專業領域。

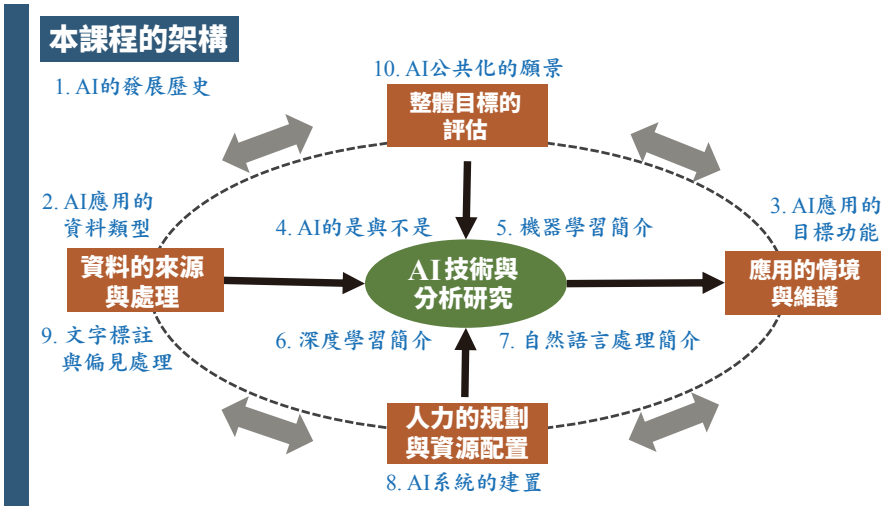
王教授先和大家分享他跨界到 AI 的經驗，從天文研究的超新星和出生恆星開始接觸到大數據的應用；接著與腦科學的教授們一起使用大腦資料庫解碼果蠅大腦的訊息傳遞；後來走到多體物理與量子電腦的物理研究；再是親權判決

[3] Linguistic Inquiry and Word Count，語文特性分析的電腦工具。

預測的司法與 AI 大數據研究，以及第三場次提及的網路自我傷害文字分析；最後是推動 AI 的公共化。冀望以自身的經歷鼓勵大家多跨出自身的專業領域，了解其他領域專家的研究，如此可以找到拓展自身專業的路徑，影響的層面也能更廣泛。

之所以會開始推動 AI 導論課程，也是因為在 AI 的公共化上試圖讓更多非理工科系的學生能夠了解及使用 AI，因為大多數具有價值的資料存在於政府與公共事業部門，這些資料的整理和評估還是需要仰賴人文社會學者的專業，而且 AI 跳脫以往科技發展是以結構性的知識為基礎，在非結構化資料的處理能力（特別是文字、圖像）部分，更需要人文社會學者以新的思維來投入研究。因為 AI 在學習資料的過程是一種統計性的結果，不是必然的物理定律，少見的個案或是結構性的問題機戶沒有辦法處理，所以在 AI 的使用，如何規定使用範圍、最終的權責歸屬、賦予其價值與意義，都是人文社會學者和理工學者必須一起面對、一起思考的（王道維、林昀嫻，2020）[4]。

王教授接著介紹 AI 導論課程的設計理念，是為了讓學生不盲目引用 AI 的新聞，能多了解 AI 的基礎技術面向，進而對 AI 當前的發展提出正確的問題，甚或利用手邊的資料來做 AI 應用，但並非需要自己親自來寫程式做 AI 計算。課程的內容共有 10 種[5]，內容與課程架構一併呈現如圖一。



圖一：AI 導論課程設計理念及內容（圖片來源：王道維教授演講投影片）

[4] 王道維、林昀嫻（2020）。〈如何用 AI 創造社會共善？——AI 公共化的契機〉，國立清華大學通識教育中心。取自 http://cge.nthu.edu.tw/cgenews147-2/。

[5] 相關影片可至 YouTube 搜尋「科技部人文社會科學研究中心」，共有十集循序漸進由 AI 發展歷史、資料與技術面向，以及各種人文社會相關範例，進而討論人文社會推動公共化 AI 的展望。

課程的應用可以依據個人的興趣隨意挑選主題影片，或僅看投影片[6]，研究增能的部分對於有興趣從事 AI 相關的人文社會學者，可以搭配閱讀相關已發表的論文或書籍，更進一步可以與技術團隊討論可能使用的 AI 技術與其資料型態、標註[7]方式、應用場域，評估研究所需要的配套資源。王教授最後提及 AI 發展日新月異，對社會各個層面影響甚深，設計 AI 導論課程就是希望能幫助人文社會的學生更了解相關的技術，以便做更好的掌握與分析，如此才不至於讓 AI 的出現造成不良的影響。

[6] 投影片可至以下網站下載：https://drive.google.com/drive/folders/1bh3jtXI1ZXmxaJlUuOPcfbDCTH8pUttt。

[7] 標註是將原始資料做結構化的整理，使之能夠在 AI 應用時產生有意義的連結與關係，對 AI 應用成效影響很大。