

# 實現公共化 AI 的未來<sup>#</sup>

林文源\*

## 一、AI 作為危機與轉機

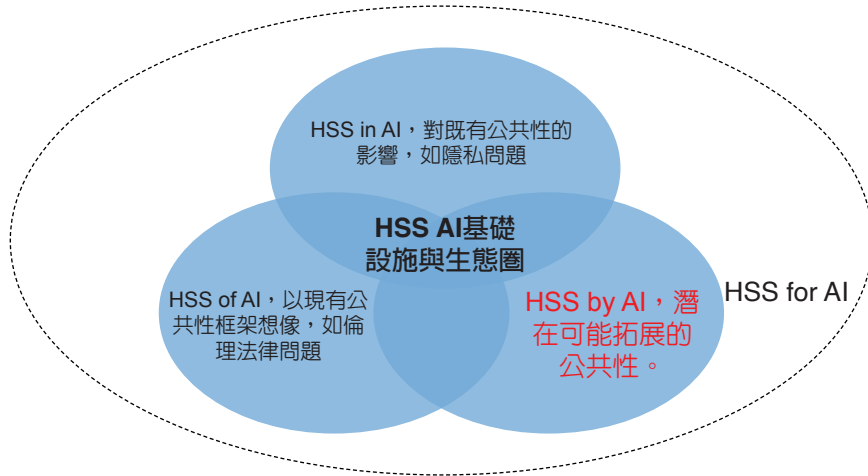
近十年來人工智能 (AI) 在全球的發展，尤其是 2022 年末起的生成式 AI (Generative AI) 風潮，對各地社會產生更深遠影響。在臺灣，由技術和資本主導的 AI 發展也引起對於社會公共性的關切。國科會 (前科技部) 不但宣告 2017 年為 AI 元年，啟動了四大中心計畫，陸續頒布在地科技研究、公務機關的 AI 相關指引，並積極推動以正體中文為主的「可信任人工智慧對話引擎」(Trustworthy AI Dialog Engine, TAIDE) 計畫。與此同時，國科會人文處於 2021 年啟動的「公共化 AI」增能專案，也透過結合人文社會和技術協作，不只關注 AI 技術轉型、倫理和監控問題，更積極探索人文社會科學 (HSS) 領域如何參與 AI 的變局與機會。大方向來說，如圖一所示，推動公共化 AI 不但希望由人文社會的價值與觀點分析與面對 AI 浪潮，也希望 AI 能有助於更廣闊的人社發展，以及以人社參與實現更好的 AI 願景 (林文源，2019)。

這些願景與工作的實現並非一蹴可幾，需要透過多方合作推動。為此，第一期的公共化 AI 增能方案透過一年方案推動奠定基礎。第二期更進一步藉由社

<sup>#</sup> 本文介紹公共化 AI 成果與展望，這些成果改寫自第一期與第二期成果報告濃縮而成，但因篇幅所限，許多討論精簡自作者既有討論，也省略許多引用，請參閱相關報告與論文。第一期「打造公共化 AI：人文社會跨領域 AI 資料中心推動」，執行期間為 2020 年 11 月 1 日至 2021 年 10 月 31 日，由國立清華大學通識教育中心林文源教授擔任計畫召集人。共同召集人則有：中央研究院李建良特聘研究員、國立清華大學物理學系王道維教授、國立政治大學公共行政學系杜文苓教授。第二期「公共化 AI (II)：朝向公共化的生態圈」計畫。執行期間為 2021 年 11 月 1 日至 2023 年 12 月 31 日，由林文源擔任計畫召集人。共同召集人有：中央研究院李建良特聘研究員、中央研究院劉紹華研究員、國立清華大學物理學系王道維教授、國立政治大學公共行政學系杜文苓教授、國立政治大學財政學系連賢明教授、國立政治大學心理學系游琇婷副教授、國立臺灣大學政治學系黃旻華教授、國立雲林科技大學數位媒體設計系楊晰勛副教授、國立陽明交通大學科技與社會研究所彭松嶽副教授。在前二期基礎上，目前正在執行第三期「公共化 AI：人社跨領域的落地應用及問題探索」計畫，執行期間為 2024 年 1 月 1 日至 2025 年 12 月 31 日，由林文源擔任計畫召集人。共同召集人為國立清華大學物理學系王道維教授。目前正持續召集以各種形式參與推動的成員。

\* 國立清華大學通識教育中心教授、「公共化 AI」專案主持人

會、STS、法律、人類、經濟、心理、政治、區域、藝術等九學門的支持持續推動。第三期則在這些基本觀念推廣、跨領域對話與合作、資料標註基礎設施上，做進一步推展。



圖一：人文社會與AI發展的公共化機會

## 二、成果

不同 AI 技術雖然各有不同重點與程序，大多包括如何框架問題（problem framing）、訓練資料、機器學習，以及詮釋、應用與持續調教（例如以人類回饋的強化式學習）等重要實作面向<sup>1</sup>。因此需要跨領域合作、擴展問題意識，並透過脈絡化的整合資料、調校更為適用與精確的演算法。然而，目前 AI 技術多由歐美巨型公司發展，尤其是現有大型基礎語言模型只包含少數正體中文資料的情形下，臺灣如何融入？在地資料在何處？有哪些資料庫或方向值得投注探索？都是大問題，相對而言，如何讓熟悉這些資料的相關人文社會領域專家參與此 AI 願景，不只需要人文社會社群集體參與，更需要前瞻規劃以形成長遠戰略。

在此願景下，公共化 AI 結合人社眾多學門召集人與代表推動者之視野推動下，本計畫已累積以下成果。

### （一）人文社會 AI 文字資料標註系統

有鑑於 AI 發展必須的資料標註工作具有根本重要性，因此在前兩期我們努力開發出開放公共使用的「人文社會 AI 文字資料標註平台」，並製作相關使用說

<sup>1</sup> 請見林文源(2023)的討論。

明與影片。顯而易見，從問題框架、資料準備與處理、標註與應用都不只是技術問題，面對人類社會問題，若要達到高於人類處理效率與精準度，必定需要提供人文社會專業觀點標註的資料，協助機器學習，以發展能落實於在地制度與文化環境的 AI。因此，如何進入具體專業領域合作推動，如何更為審慎考量不同資料、價值與理念，且有效率的吸引參與協作，是推動人社 AI 方案的精髓。藉由推廣標註系統與吸引更多接觸，將大幅降低人文社會領域要投入的門檻與資源需求。因此，我們首先致力於建立公共化的文字標註平台，協助本地人文社會學者將文字轉換為結構化資料，作為 AI 應用的訓練資料集，同時可供其他量化研究使用。該平台目前托管在國立清華大學人文社會 AI 應用與發展研究中心（清大人社 AI 中心）網頁上，建置完成後以符合公共 AI 理念發布，公開接受試用。<sup>2</sup>

## （二）跨領域媒合

在跨域協作上，我們也成功舉辦九場跨領域論壇<sup>3</sup>和十場學門工作坊。來自各校和各領域的參與者，不但皆為有志者或有經驗者，也有不少正積極探索各種公共化 AI 的推動，都有助於瞭解各學界生態與共識<sup>4</sup>。這一階段不僅旨在推廣人社領域的資料和問題意識有豐富收穫，也成功媒合幾位技術相關的合作夥伴。希望未來能促成更多推廣和合作範例，匯集有志於公共化的團隊。

本面向成果主要體現在四大主軸，都與完善公共化 AI 的生態圈相關：首先，成功培養人社團隊對 AI 發展基礎的具體理解，包括相關 AI 知識和具體細節討論，不僅僅只停留在對 AI 未來的擔憂或以學科視野分析 AI。其次，我們厚實跨界視野。投入 AI 研究與發展需要跨足多個領域，每一次合作都是一個獨特的跨領域模式，需要人文社會領域的專業知識，同時對相關 AI 技術有所瞭解。由於人文領域和 AI 技術領域的資深學者受限於自身的訓練和視野，因此這種跨領域計畫更適合具備人文社會領域博士後研究或年輕學者的專案管理人才。這有助於將領域學生的生涯規劃由既有領域擴展到相關的 AI 應用與影響評估，這需要經驗累積和合作經驗，我們在每次工作都持續在嘗試示範與落實這些面向。其三是強調了培訓人社團隊中的 AI 專案人才的重要性。這種人才在資工教育訓練較不受重視，但卻難以由人社訓練發展。這些人才必須具備跨資訊背景，並對領域知識和資料狀態有一定的瞭解，才能逐漸培養。特別是人社領域的人才可以通過資源轉介和輔導過程培訓，使其將領域中的資料轉換為 AI 可

<sup>2</sup> 請見：<https://nthuhssai.site.nthu.edu.tw/p/404-1535-243559.php>，並請參考相關使用文件與操作影片。

<sup>3</sup> 九場跨領域論壇成果請見：<https://nthuhssai.site.nthu.edu.tw/p/403-1535-9495.php>。

<sup>4</sup> 十場公共化 AI 學門工作坊成果請見：<https://nthuhssai.site.nthu.edu.tw/p/403-1535-9859.php>。

以學習的內容。最後，逐漸開發相關種子資料和應用的團隊。這也是目前在跨領域推動中最需要實現的，但卻並未有相關的正視管道鼓勵。我們希望後續能持續擴大接觸與醞釀交流的 AI 生態圈，將這種開發機制擴展，在各人社領域形成資料開發的機制，這樣才有助於將現有越來越多的需求、技術、資料與制度轉為公共化 AI 的基礎。

### (三) 知識公共化

在先前推動過程，因為體認到以個別學者摸索有其人力物力限制，為了加速理念與合作模式之擴散，前兩期已經進行人社 AI 教學叢書之規劃，並編輯出版兩本人文社會 AI 跨領域專書，以及完成線上「人社 AI 導論」課程十集和公眾推廣文字與影音數十項，希望此號召有志者。其中包括：

1. **AI、科技與社會系列叢書：**與清大出版社合作，設立臺灣第一套「AI、科技與社會」系列叢書，由林文源擔任主編，邀請國內重要人社 AI 相關重要學者李建良、杜文苓、吳重禮、吳齊殷、張卿卿、蔡政宏等學者擔任編輯委員。採公開徵稿方式，共同為人文社會 AI 相關技術研究、教學與推廣努力，希望能提供一般社會大眾及人文社會與 AI 教科用書，創造更適合本地發展之公共化 AI 發展。<sup>5</sup>
2. **跨領域專書：**推動 AI 公共化的過程中，我們參與並推動兩本書籍的出版，記錄當前臺灣少見的跨域合作經驗，希望啟發更多人社 AI 跨領域嘗試。第一本專書為：李建良、林文源編（2022），《人文社會的跨領域 AI 探索》，國立清華大學出版社。第二本專書：林文源、王道維、杜文苓、李建良編（2024），《公共化 AI：思維、協作與法制的基礎設施》，現正出版中。
3. **公共推廣：**推動 AI 公共化需要各方參與，包括 AI 適用性定義、資料處理、關鍵演算法發展以及場域建置等考量與影響的討論。為擴大認知，我們致力於積極參與演講、投書，以及參與公共機構的研討等形式，不斷積累公眾對於 AI 技術、風險、政策和社會影響等公共議題的理解，期望促進更廣泛的討論，進而擴大對 AI 公共性的想像與認知。<sup>6</sup>

## 三、公共化的未來：挑戰與策略

當然，這些過程也非一帆風順，目前我們也同步意識到不少挑戰，尤其穩固 AI 基礎設施的議題。一般來說，基礎設施往往指的是硬體設備，但在 AI 發

<sup>5</sup> 請見出版社網頁：<https://thup.site.nthu.edu.tw/p/404-1210-240758.php?Lang=zh-tw>。

<sup>6</sup> 請見：<https://nthuhssai.site.nthu.edu.tw/p/404-1535-247400.php>。

展業無可避免更多「軟」性基礎設施包括資料等，例如臺灣 AI 量能評估報告 (Taiwan AI-Readiness Assessment) 的 AI 基礎設施指的是算力 (computing power) 與資料 (data) 基礎設施。但根據上述對於 AI 實作介紹，應當能夠體會除了硬體外，還有諸多牽涉在地特質的議題。當前最大的挑戰來自在地資料、跨領域合作、人社學界對於 AI 的共事與合作，與生成式 AI 的新挑戰幾方面：為此，當前目標著重在「落地」實現，希望形成更為整體與永續之動能。這包括：

### (一) 人社領域之生成式 AI 指令工程 (AIGC Prompt Engineering for HSS)

自從 ChatGPT 等生成式 AI 在 2022 年底問世以來，這些大型語言模型 (Large Language Models) 對人文社會領域的各個層面都帶來新的影響，尤其是以語言互動的指令 (Prompt) 溝通方式，讓這些生成式 AI 更迅速普及。雖然坊間已經出版供許多領域應用的指令書籍或課程，但是整體而言卻不足以應用於對於文字精準性或思辨性要求較高的人文社會領域學習，但卻對人社研究與教學帶來了重大挑戰與機會，為因應這種變化，本計畫在 2023 年初進行初步討論，並於 6 月正式啟動「生成式 AI 指令工程培訓計畫」。在 2023 年 9 月公布「人社課程用生成式 AI 指令範例：通用版」<sup>7</sup>、「人社課程用生成式 AI 指令範例：邏輯與論證示誤版」<sup>8</sup>，以及「人社課程用生成式 AI 指令範例：科技與社會」<sup>9</sup>，迄今有一萬七千多次下載，顯示這些生成式 AI 與這些指令集受到人社學界的廣泛重視與需求。目前第二波有新的數門種子課程指令集也就緒，即將開放交流參考。

### (二) 公共化種子資料集與研究範例

人文社會的特色與重大價值在於多元與百花齊放。如上關於 AI 協作與拓展公共性的討論顯示，AI 發展不能只由硬體與技術主導，最終還是要以需求（及資料）為王，在「數據驅動」(Data-driven) 及「問題解決驅動」(Solution-driven) 思維下，就不能只是移植或是應用既有 AI 技術，應當由人文社會的多元視野，匯集在地脈絡的需求與資料「長出」適當的 AI 問題與發展方向。總而言之，資料、硬體與技術等基礎設施需要同時相互磨合推進形成永續生態圈。

<sup>7</sup> 請見：[https://www.dropbox.com/scl/fi/b1czzntb37qtl11wzljji/Prompt-template\\_.docx?rlkey=ry1in6rtu9nudl64tctfx7se&dl=0](https://www.dropbox.com/scl/fi/b1czzntb37qtl11wzljji/Prompt-template_.docx?rlkey=ry1in6rtu9nudl64tctfx7se&dl=0)。

<sup>8</sup> 請見：[https://www.dropbox.com/scl/fi/f6zllhqaohu340mvtj8ap/Prompt-template\\_.docx?rlkey=92576bdgp1apkee14u92rwl2&dl=0](https://www.dropbox.com/scl/fi/f6zllhqaohu340mvtj8ap/Prompt-template_.docx?rlkey=92576bdgp1apkee14u92rwl2&dl=0)。

<sup>9</sup> 請見：<https://www.dropbox.com/scl/fi/jdfs9v3f41j3jie7vsnik/20230904-ST5-AIGC.docx?rlkey=8zfl0c1dv6uc1ps8we1jxn2to&dl=0>。

儘管當前已經有許多生成式大型語言模型出現，反而代表更需要在地文字資料與標註。在大型模型用途越來越廣時，這些少量經過專家標註過的人工標註資料，反而是更為寶貴可以用來精準回應在地需求的基礎。所以累積本地標註資料集，以及開發在地文字標註平台與應用，在現階段仍有其不可取代的重要性。因此，本計畫未來也將積極與成熟之有志團隊合作，進一步整理資料並將之公開為「種子資料集」。

### (三)自動標註作為公共 AI 賦權平台基礎設施 (Civic AI Capacity Infrastructure)

如上成果所述，有鑑於 AI 發展必須的資料標註工作具有根本重要性，因此在前兩期我們努力開發出開放公共使用的「人文社會 AI 文字資料標註平台」，並製作相關使用說明與影片。這大幅降低人文社會領域要投入的門檻與資源需求。這一期與資料集累積相關的標註工作上，我們將嘗試進一步發展自動標註。現有國內相關 AI 教育訓練甚少使用人社領域的資料，即使有已是個別團隊累積，很難深入擴大或累積，主要因素為各領域資料差異甚大，且多半並未花額外力氣整理公開各界使用，因此需要有各領域的種子團隊開始落實。若有種子資料之後，要能擴大標註的資料量是不可少的，其中方向之一是使用 AI 來作自動標註的工作，未來這部分的工作的重要目標是在理據 (Case Based Reasoning) 標註平台基礎上建立自動標註。

### (四)AI 落地過程之人社倫理共識

當 AI 持續牽動社會發展，相關公共化應用與共識也必須隨之調整，這也包括研究領域。因此，人社 AI 之發展與應用，必須關照各地差異外，也需逐漸凝聚在地研究展望等規範性基礎設施。因此，未來我們也希望能逐步朝向由下而上、形成關於在地相關 AI 研究規範與展望之共識，尤其是在權衡如何由社會與技術安排共同協作，在哪些適當條件下，以避免傷害與不當影響為前提，鼓勵人文社會領域參與共創更多 AI 的可能性。

我們期盼由這些方向持續推進，能有助於更為多元的公共化 AI 的未來。

## 參考文獻

- 林文源 (2019)。〈AI 的能與不能：以醫療與照護為例發揮人社想像力〉，《人文與社會科學簡訊》20 卷 2 期，頁 99-103。
- 林文源 (2023)。〈STS 如何介入 AI：ANT 取徑朝向 AI 公共性的思考實驗〉，《科技、醫療與社會》36 期，頁 233-270。