

「知識之眼」研討會全紀錄： 以 AI 探索臺灣人文的新篇章

葉素玲*

在資訊如海的今天，要如何探尋歷史的真相？如果想研究一段特殊的史料，往往意味著要在浩瀚的文獻中艱辛探索，不僅耗時費力，也常被零散或不實的資訊所困擾。如今，人工智慧 (Artificial Intelligence, AI) 的崛起，彷彿為我們創造了一雙強大的「知識之眼」。這項技術以前所未有的方式快速整合資訊，為傳統的研究方法帶來了突破的契機。

然而這雙新生的眼睛並非完美，它有時會出現「幻覺」(Hallucination)——在缺乏根據時，產生看似合理卻與事實不符的內容。這對於講求精確證據的人文研究來說是一大隱憂。我們該如何校準與駕馭，讓它成為可靠的研究夥伴，而非混淆視聽的來源？



圖一：「知識之眼」研討會講者於臺灣大學博理館合影

* 國立臺灣大學心理學系暨研究所特聘教授

為此，一場以「知識之眼：AI 探索臺灣人文——挑戰與未來」為名的研討會於 2025 年 6 月 14 日在臺灣大學博理館國際會議廳舉行。這場活動邀請了 AI 技術與人文領域的專家學者，提供一個跨學科的交流平臺，以新的視角與方法，共同探討如何運用 AI 的新視角與方法來推動臺灣人文研究的發展。

一、第一部分：一個為臺灣而生的 AI——「臺鵠」系統的設計與誕生

研討會的上半場，焦點集中在一個專為臺灣人文研究量身打造的 AI 系統——「臺鵠」(TAIHU, TAIwan Humanities Conversational AI Knowledge Discovery System)。其中「鵠」字取「鴻鵠」之意，展現了在臺灣人文知識領域探索的目標，並致力於建置生成對話式的知識探勘平臺。它的開發歷程，清晰地展現了要打造一雙精準、可信的「知識之眼」所付出的努力。

(一)場次一：從問題到解方——臺鵠系統的開發願景

研討會由我(臺鵠計畫主持人)揭開序幕。首先點出了人文學者面臨的兩大挑戰：珍貴的歷史文獻資料分散各處不易檢索以及通用的大型語言模型時常產生「幻覺」，對於需要嚴謹考證的學術工作而言並非理想的工具。

為了解決這些問題，臺鵠計畫應運而生。團隊的核心策略是採用「檢索增強生成」(Retrieval-Augmented Generation, RAG) 技術。我們可以將 RAG 想像成一個正在進行「開卷考試」的 AI。它在回答問題前不會憑空作答，而是會先去一個指定可信的「圖書館」(也就是專業資料庫)中查找相關資料，然後才根據這些查到的資料來生成答案，這個機制從根本上提升了回答的準確性與可信度。

更重要的是臺鵠的圖書館是獨一無二的，它整合了多個臺灣特有的人文資料庫，例如《臺灣省議會公報》、《臺灣明清行政檔案》、《國家文化記憶庫》以及《漢文臺灣日日新報》等，並陸續有更多授權資料庫在建置中。這意味著臺鵠的知識根基，是深植於臺灣本土的歷史脈絡。為了確保這雙眼睛看得準，團隊更從零開始，建立了超過 600 筆的評測標準 (Benchmark) 題組，並結合事實查核系統與使用者經驗研究，以確保系統能真正滿足學者的需求。

Technology is best when it brings people together，期許臺鵠系統能成為凝聚學界力量的橋梁，推動臺灣人文研究與 AI 深度整合。透過跨領域的協作與資料授權的突破，持續擴充知識邊界，未來甚至能從被動的資料查詢，朝協助學者「形成研究問題」等更高層次的需求邁進。

(二)場次二：實戰檢驗——臺鵠與 ChatGPT 的比較分析

有了理想的設計，實際表現如何？臺灣師範大學臺灣史研究所的張素玢教授與臺灣大學歷史學系的陳冠妃教授，以具體的臺灣史問題，對臺鵠與 ChatGPT 進行了測試。

臺鵠的開發資源遠遠不及 ChatGPT，但它的設計原則卻非常明確。面對問題，臺鵠的回答篇幅相對簡短，但字字句句都嚴格依據其串接的資料庫內容，並能附上可追溯的文獻出處與原文連結。它的資訊準確性與可信度極高，當被問及資料庫未涵蓋的問題時，臺鵠會誠實地回答「無資料」，刻意抑制任何無依據的生成行為。

相比之下，ChatGPT 能生成篇幅更長、文筆流暢的答案，即使沒有直接的資料支持，它也會整合網路上所有可得的資訊，進行推論式回應。這種能力使其用途廣泛，但也帶來了學術應用上的巨大風險——這些推論很可能包含虛構或不實的「幻覺」內容。

這場比較清晰地揭示了兩種截然不同的設計邏輯，ChatGPT 追求的是通用性與流暢的對話體驗，而臺鵠則選擇了一條對學術研究至關重要的道路：犧牲了部分靈活性與生成篇幅，以換取資訊的絕對可驗證性與信賴基礎。對於需要精確引證的學者而言，臺鵠這種「知之為知之，不知為不知」的設計是更值得信任的研究工具。

(三)場次三：奠定基石——從零開始的評測之路

要讓 AI 學會看懂人文問題，首先需要一套客觀的「評量尺」，也就是標準化的測試資料集。中央大學資訊工程學系的蔡宗翰教授分享了建置臺鵠 Benchmark 測試資料集的過程。他指出，過去臺灣人文領域完全缺乏可用於評估對話式 AI 的標準化測試資料，這意味著團隊必須從零開始。

這不僅需要重新定義評測任務、設計多樣化的問答類型，還必須針對多個來源的資料庫，耗費大量人力與專業判斷，去編寫對應的標準答案與評分準則。蔡教授特別強調，人文問題本身在語意上的模糊性與開放性，為 Benchmark 的設計帶來了極高的挑戰。一個歷史事件的詮釋，可能因不同史觀而有多種答案，這與非黑即白的數理問題截然不同。

這項標準化測試工作為臺灣人文領域的 AI 對話系統品質評估提供了方法，其成果不僅能作為臺鵠系統的優化依據，也能供未來相關研究參考。

二、第二部分：以人為本的思考——AI 時代的挑戰、機遇、與倫理

當技術的輪廓逐漸清晰，更深層次的哲學與倫理問題便浮上水面。研討會的第二部分，學者們將視角從「如何建構 AI」轉向「如何與 AI 共處」，深入探討了這項新的技術可能將如何改變學術、教育，甚至是我們本身對知識的看法。

(一) 場次四與場次八：跨域思辨——人文學者在 AI 時代下的叩問

研討會的兩場綜合座談匯集了來自歷史、語言學、教育心理、哲學與數位人文等多個領域的專家，進行了深入的跨域討論。第一場座談由國科會人文處的黃俊儒處長主持，與談人包括李仁淵教授（中央研究院歷史語言研究所）、胡其瑞教授（國立彰化師範大學歷史學研究所）、宋曜廷教授（國立臺灣師範大學教育心理與輔導學系）與謝舒凱教授（國立臺灣大學語言學研究所）；第二場則由臺灣大學資訊工程學系的項潔教授主持，邀請到李卓穎教授（國立清華大學歷史學研究所）、岳修平教授（國立臺灣大學生物產業傳播暨發展學系與心理學系）與陳登武教授（國立臺灣師範大學歷史學系）共同對談。

座談中的討論提到 AI 不僅是提升效率的工具，其影響力也擴及學術實踐與敘事權的重塑，因此需要各領域學者主動積極參與。在深入的討論中，與會者就多個關鍵議題交換了意見。在教育方面，當 AI 能輕易生成文本時，「原創性」



圖二：由黃俊儒處長主持的綜合座談，聚焦 AI 為人文學界帶來的機遇與挑戰（圖片由左至右為謝舒凱教授、宋曜廷教授、黃俊儒處長、李仁淵教授、胡其瑞教授）

的概念正被重新定義。未來的教育關鍵，不再要求學生從零寫作，而是引導他們發展出更強的批判、詮釋、與整合能力，學會如何提出好問題，並對 AI 生成的內容進行有深度的再創造。

AI 的普及帶來了所謂「方法論的民主化」，過去許多複雜的數據分析或文本探勘方法，只有少數受過專門訓練的專家學者能夠掌握。而 AI 的出現讓這些複雜的分析過程變得較為容易，這也讓學者本身獨到的詮釋能力顯得更加重要。

學者們也對於「知識殖民」提出了深刻的警示，如果我們全盤擁抱由全球科技巨頭主導的 AI 模型，其背後的資料與價值觀可能會不自覺地侵蝕當地的文化，形成一種新的文化殖民。因此，發展如臺鵠這樣具有在地文化視野與資料基礎的系統，並在教育與制度層面推動跨域人才的培育變得至關重要。我們需要培育兼具人文素養與科技能力的人才，以確保技術的發展與維繫我們自身的文化主體性。



圖三：由項潔教授（右一）主持的第二場綜合座談，探索 AI 在臺灣人文研究的未來圖景（圖片由左至右為陳登武教授、李卓穎教授、岳修平教授）

（二）場次五：馴服幻影——深入「幻覺」問題的核心

除了人文哲思，技術專家也給出了應對 AI 風險的具體策略。臺灣大學電機工程學系與資訊工程學系的傅立成教授，便針對前述不斷被提及的「幻覺」問題進行了更深入的剖析。

傅教授指出此現象意指模型在缺乏真實依據時，仍生成看似合理、實則虛構的內容。他從語言模型的生成邏輯出發，解釋了幻覺的成因，並系統性地介紹了當前應對此問題的幾種策略工具，例如，可以採用結合外部知識檢索的技術（如 RAG），讓 AI 在回答前先查證可靠資料，以有效抑制幻覺。此外，也可以運用語意相似度比對與可信度評估等機制，透過演算法比對生成內容與原始資料，藉此辨識並修正可能的錯誤。

三、第三部分：眺望地平線——AI 的落地應用與未來互動

研討會の後幾場演講，將目光投向了更廣闊的應用場景與更具前瞻性的互動模式。這雙「未來之眼」，不僅能回望過去，更將深刻地融入我們生活的各個層面。

（一）場次六：從理論到實踐——金融業的 AI 落地啟示

為了讓 AI 的應用更加具體，臺灣大學資訊工程學系的張智星教授，以金融領域為例，分享了生成式 AI 從理論走向實踐的經驗與挑戰。金融業是一個高度重視合規與風險管理的領域，對新興技術的導入特別謹慎，這使其成為一個絕佳的觀察案例。

張教授指出，生成式 AI 在金融業的客戶服務、風險分析、合約解讀與報告自動生成等方面，已展現出顯著的效率提升與創新潛力。他透過具體案例說明，AI 模型能如何協助金融從業人員簡化繁瑣的資料處理流程，強化對市場資訊的判讀能力並提升與客戶的互動品質。

然而，新技術落地也面對的制度性挑戰，包括嚴格的合規規範、法律責任的歸屬，以及客戶資料的保密問題。這些問題的解決，不能單靠技術突破，而必須仰賴跨領域的專業知識整合、持續的員工教育訓練，以及相應的制度與法規調整。這場演講為 AI 在其他專業領域的應用提供了一幅清晰的藍圖，提醒我們在擁抱技術創新的同時，制度的配套與倫理的考量也必須同步跟上。

（二）場次七：超越關鍵字——未來 AI 研究助手的樣貌

臺灣大學電機工程學系的李宏毅教授進一步探討了 AI 應用未來的趨勢，即從傳統的關鍵字搜尋，朝向語音與自然互動發展。李教授介紹了多項技術應用實例，包括語者模仿、情緒調整、與多語言生成等。這些技術展示了 AI 模型已能模擬人類對話，並具備一定程度的語境理解能力。隨著模型進一步發展，未

來的 AI 研究系統可能從被動的搜尋引擎，轉變為能與研究者進行深度對話、理解提問脈絡並提供針對性回應的協作者。

李教授也提醒，這類技術伴隨著倫理與安全上的考量。如語音偽造與身分模擬等技術，可能帶來詐騙或假訊息的風險。為了確保技術兼具可控性，呼籲開發者應在追求創新之餘，應納入更多監督與驗證機制。



圖四：活動結束後全體與會來賓合影

四、結語：回望的啟示

本次研討會不僅介紹了臺灣人文對話式 AI 臺鵠系統的初衷與目前成果，更匯聚了科技與人文領域的專家學者，促成了深度跨領域的對談。透過這場交流，與會的人文學者得以更了解 AI 的潛力與限制，技術專家也對人文領域的需求與考量有了更深入的認識。

回歸最初的主旨，這雙「知識之眼」並非一個被動觀看的冰冷鏡頭，它也能是一個主動、充滿潛力與可能性的協作者。建置與駕馭它的過程，實際上也是我們重新審視自身知識體系、研究方法、與文化價值的過程。

要讓這雙眼睛看得更清晰、看得更有智慧，最佳的方法就是讓科技的邏輯與人文的關懷緊密結合，讓代碼的建構與文化的詮釋攜手前行。臺鵠的建置是個起步，唯有攜手同行，我們才可能走得更遠！

知識之眼：AI探索臺灣人文

國立臺灣大學 博理館 101會議室 ——挑戰與未來

2025.06.14 (六) 8:30 - 17:10



**AI視野下的
臺灣人文研究新契機**
9:00 - 9:30
葉素玲



**洞見臺灣—
臺語與主流生成式AI
的實戰對話**
9:30 - 10:00
張素芬 陳冠妃



**從蒐集到搜尋：
人文生成式AI建
置的挑戰與對策**
10:10 - 10:40
蔡宗翰



綜合座談：人生成式AI的挑戰與機遇
10:40 - 12:00
黃俊儒 李仁淵 宋曜廷 胡其瑞 謝舒凱



**真偽之辨：
生成式AI結果的
事實查核技術與方法**
14:00 - 14:30
傅立成



**從理論到實踐：
生成式AI於金融的落地與挑戰**
14:30 - 15:00
張智星



**超越關鍵字：
未來AI研究助手
的語音與
自然互動趨勢**
15:10 - 15:40
李宏毅



綜合座談：探索應用AI於臺灣人文研究的未來圖景
15:40 - 17:00
項潔 李卓穎 岳修平 陳登武 張隆志



TAIHU
Taiwan Humanities Conversational AI
Knowledge Discovery System



NSTC 國家科學及技術委員會
National Science and Technology Council

報名連結：

02-3366-3955
taihucais@gmail.com

主辦單位：國家科學及技術委員會資助之「生成式對話之臺灣人文知識探勘系統計畫」（臺語）
協辦單位：國立臺灣大學數位人文研究中心、國立臺灣大學人工智慧與機器人研究中心
國立臺灣大學心理學系暨研究所

圖五：「知識之眼」研討會海報