AI於高等教育品質保證評鑑之導入與反思

文／高宏宇．國立清華大學資訊工程學系教授

科技與人文長久以來被視為兩股拉扯的力量，科技追求效率、可計算性與邏輯性。而人文則重視價值思辨，倫理判準與文化脈絡。尤其在高等教育領域，這兩者的關係更顯複雜。在追求科技進步與數位轉型的浪潮下，高教可能走向技術至上，人文式微的危機。然而，這也激發了對「科技與人文融合」的新思維。特別是人工智慧與大數據分析，確實能夠為高教帶來突破性的創新，如智慧課程推薦、學習歷程分析、教學資源優化等。但這些科技若無人文精神的引導，極可能淪為技術工具的自我繁殖，缺乏對教育價值與倫理原則的反思。

以生成式人工智慧（Generative artificial intelligence，簡稱生成式AI）為例，大型語言模型（LLM）如GPT系列、LLaMA、Gemma等，具備強大的自然語言理解與生成能力，能在短時間內產出大量語句通順、邏輯清晰的文字內容。這對於教育行政、學術研究乃至品質保證等工作皆提供新契機。但我們不能忽視這些語言模型背後的語料來源、訓練偏差與社會語境可能深深影響輸出結果，其倫理風險與透明度議題是需要人文學科介入協助釐清。

也因為AI技術爆炸性的突破，大型語言模型的成熟且落地應用比比皆是，利用這樣技術與工具輔助與進行大學品保評鑑成為學校與品保評鑑單位一項課題。作為一位研究自然語言處理以及生成式AI大型語言模型技術的學者，在導入這項技術所帶來的應用潛能外，也正視到可能衍生的社會風險與教育衝擊。以下從技術原理、演化趨勢與實務應用等層面，討論AI技術在高等教育與品質保證評鑑中的成敗關鍵。

AI技術具效率、擴展性與語言理解

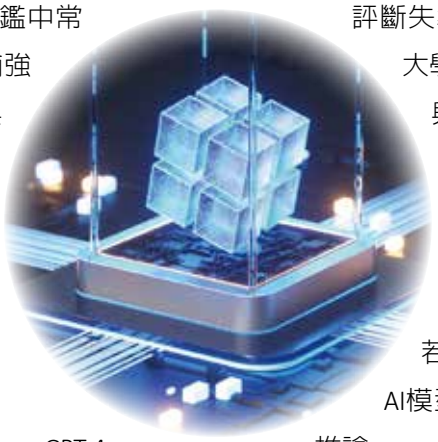
一、文本理解與摘要能力

大型語言模型具備極高的語義解析能力，能迅速掌握文件主旨與關鍵句，在教學評鑑、審查報告、政策文件中可以產生極高的摘要與分類準確性，減輕人工負擔。例如，我們透過國內架構在LLaMA模型上的TAIDE與Breeze模型微調訓練與加上檢索強化生成（Retrieval-augmented generation, RAG）的審查系統，可準確辨識教育目標對應、教學方法與評量一致性等問題。並可在多文本中進行資訊消化與理解，讓數據為重的教育品保評鑑能夠靠AI更快速更精準地整理出審查需要的學校量化數據與質化表現。

二、知識生成與重構

LLM具備將多來源、跨年度資料重組、推理與語言化的能力，能從片段不完整的資料中還原

出概念意涵的脈絡。這對教育評鑑中常見的資料片段化現象是很好的補強工具。以往評鑑人員利用檢索與查詢來收集片段化的資訊，利用人工整併以及推理後加以論述。現在這流程可直接讓AI模型進行自動化整併，可提高資訊完整度。



三、語言跨域轉換能力

多語言訓練的語言模型如Gemma、GPT-4o已能進行具文化敏感度的翻譯與文本改寫，能大幅提升非英語母語地區大學於國際評鑑中的文件參與能力，實現教育國際化參與門檻的降低。

四、可擴展性與模組化訓練

以LoRA（Low-Rank Adaptation）、Adapter等參數高效微調技術為基礎，AI系統可針對特定校務資料進行輕量化在地化調整，只要提供在地校務資料就能提升模型的專業性與回應能力，同時避免全面再訓練帶來的成本與資源壓力，且通常也伴隨著破壞原始模型能力的現象。

五、即時互動與視覺化整合

搭配前端儀表板與Chat介面，語言模型可協助評鑑人員與學術單位建立問答式互動平台，即時回應校務治理相關查詢，例如：「哪些系所在近五年內修習雙學位學生增加？」或「學院在產學合作上是否已達成計畫目標？」

AI技術風險與限制：語料偏誤、理解錯誤與倫理失衡

一、語料偏誤影響公平性

大型語言模型的基礎訓練語料多來自網路開放資料，其語言使用偏向英語世界、科技產業與主流價值觀。這使得在處理地區型、非主流文化背景的大學資料時，模型可能因語言風格不熟悉而評斷失準。例如在原住民族研究或宗教大學的報告審查中，模型對敘事邏輯與文化對齊（alignment）上經常理解錯誤。

二、資料品質與標準化問題

高等教育領域的報告格式、敘事習慣、術語用法差異極大，若無完善的前處理與語料標準化，AI模型容易誤解文本語意，導致錯誤推論。

三、一般性常識的不足與概念偏移

語料偏差也會造成對於特定文化的一般常識（common sense）理解不足，而用不足或是偏差的常識對問題產生錯誤理解以及推理，就會產生看似正確但論證邏輯卻不符事實狀況情形發生。教育評鑑牽涉多層次制度安排與文化語境，AI雖擅長語言處理，卻難以深度理解背後的政策意圖與歷史脈絡，須仰賴人工補充與驗證。

四、資料隱私與倫理風險

許多校務資料含有學生、教師個資與敏感內容，若無嚴密的資料治理規範與AI模型訓練界線，恐觸犯個資法規與倫理爭議。

五、缺乏真正的推理與事實驗證能力

現階段語言模型擅長樣式比對與對齊（pattern matching），但非真正具備邏輯推理與即時資料驗證能力。這可能導致模型給出語法上合理但邏輯上錯誤的結論。例如對於學生成就評估指標的預測，模型可能錯誤推斷出「教師升等率」與「學生學習成效」具高度相關性，進而誤導決策。

六、透明性與可解釋性問題

多數生成式AI的內部運算為黑箱機制，無法提供具邏輯層級的推理脈絡。這使得在評鑑爭議時難以舉證來源與依據，使得結論多半不可溯源也不可證明。當AI模型做出評估建議時，若無法

清楚說明其依據與邏輯，將難以取信於專業審查者，亦無法作為正式決策依據。這也限制了其在制度性品質控管上的角色。

七、模型「幻覺（hallucination）」風險

語言模型在無資料依據下生成虛構內容的問題依然存在，若不慎將模型產出的內容直接引用於正式報告，將產生學術誤導、評鑑錯判的嚴重後果。雖目前模型進步相當快，這幻覺問題也被正視與研究，但仍無法保證其正確性。現今一般採用的RAG技術，僅能降低此幻覺生成現象，並無法解決整個問題。且因生成過程充滿隨機性，在驗證與檢核上也相當不易。

八、AI取代與專業角色弱化

部分學校在預算壓力下，可能傾向過度依賴AI進行評鑑文件撰寫與審查初審，導致原本應由具教育經驗與學術敏感度之教師或專家負責的工作遭「自動化取代」，長期將稀釋教育專業的價值。

AI技術方向與成功應用可行性

綜合以上AI強項與潛在問題，大型語言模型與生成式AI技術在教育品質保證中的應用潛力巨大，但亦暗藏風險。唯有技術社群與教育工作者共同協作，才能打造出既有效率又具倫理與文化敏感度的AI系統，真正服務於高等教育的永續發展。現今導入面在技術與人因整合上也有一些努力，如：

一、往結構化與因果推理發展

未來的語言模型將逐步整合符號邏輯（symbolic reasoning）與圖結構表示（graph-based representations），使其具備更強的事實一致性與概念邏輯連結能力。應鼓勵推動「結構化教育評鑑語料」的建立，如表徵學習過的評鑑指標系統與案例知識圖譜。

二、發展「教育語料透明指標」

所有高教評鑑使用之AI模型，應具備語料來源揭露、偏誤評估報告與開放檢驗機制。透過建立「語料足跡註記」與公平性測試報告，建立可信任的技術監督框架。

三、強化跨領域訓練與研究合作

推動資訊科學與教育、人文、社會、法政等領域協同設計AI模型訓練策略，納入在地文化詞彙、評鑑語境知識、教育價值反思等維度。

四、重申人為主體的設計原則

技術應始終服務於人，尤其在教育場域，AI的功能應輔助人類思考、放大專業視野，而非取代判斷或規避責任。透過互動式使用方式，除了即時微調系統表現外，也可以讓模型回覆更貼近使用者問題，也能減少模型所產生的幻覺與虛假的回應。

AI於高等教育品質保證評鑑可能性應用與策略

高等教育品質保證涉及教學、研究、行政等面向，需透過結構化與系統化的方式，確認各項教育活動是否達成其應有的目標。在此過程中，AI尤其是生成式AI，目前已能夠發揮以下幾項關鍵功能：

一、計畫書與報告審查輔助

高等教育機構每年需撰寫大量自評報告、發展計畫、教學成果報告等，供品保機構或校內評鑑使用。LLM能夠分析文件結構，自動歸納摘要、找出關鍵指標對應程度，並提供可供參考的審查意見草稿。這不僅節省審查人力，也提升文件審閱的系統性與一致性。審查者可以更有效率找到計畫書中散落各地的相關資訊，並能透過理解與邏輯推演產生論述與問題回應。

二、文本脈絡理解與知識整合

AI能整合歷年審查紀錄，辨識特定科系或學程的發展脈絡，除自動化擷取計畫書內容外，也能

提供比較與分析。並評估其與教育部政策，與高教深耕等制度的對接程度。例如自動分析教師專業發展與課程設計的一致性，或評估學生學習成果與職場銜接程度。

三、趨勢預測與異常分析

透過多年度評鑑數據，AI可協助辨識學校在特定面向上的趨勢，如教學評量下降、學生就業率改變、或研究成果分布異常，進而提出風險預警機制，供校方與品保單位提早因應。

四、智慧型決策支援系統

透過語言模型建構的互動平台，可讓校務主管與評鑑人員進行自然語言查詢，如「本校近五年外籍生註冊率變化為何？」或「本學院是否落實多元入學精神？」AI可即時從資料庫中擷取資訊並生成文字說明，提升決策效率與知識透明度。

五、教學與評鑑語言的平衡機制

傳統的自評與報告常因敘述風格或語言表達影響審查觀感，AI可協助進行語言風格標準化處理，提升文字平衡性與中立性，避免因表達落差產生不公平的評估結果。

六、多語言與跨文化評鑑支持

AI能進行跨語言翻譯與文化轉譯，使不同文化脈絡中的大學能彼此理解評鑑準則與發展邏輯，促進國際合作評鑑平台的建置與資料互通性。

七、強化教育公平性評估

AI可協助辨識不同群體（如原住民學生、離島學校、社經弱勢學生）在教學資源與學習成果上的差距，協助大學進行校務策略修正，實踐教育公平與社會正義的核心理念。

面對AI對高教帶來的變革，大學與政府兩方在面對這技術時必須同步展開調適與治理策略，對大學端而言，應設計針對AI理解、資料倫理、模型風險識別等面向的培訓課程，讓使用者能判斷AI產出是否合理、是否需人工校正。並鼓勵各校發展結合AI的自我評估工具，將數據分析與文字理解功能整合進校務治理流程，促進品保的持續性與精準性。以及成立AI倫理委員會或治理小組，定期審視AI模型使用範圍、風險回報機制與倫理指引，確保應用不脫離教育使命與公平原則。並結合資工、語言、教育、法律等領域，共同開發AI評鑑工具，讓師生共同參與創新研發，提升應用的實際性與可接受度。

而對政府與品保機構而言，也應制定AI應用準則與技術標準，例如針對評鑑用途的語言模型，訂定其準確性下限、解釋性要求與資料來源規範，避免低品質AI干擾正式判斷。投入資源建立可跨校使用的開放型AI評鑑工具，提升中小型學校之技術可近性，並集中維運以確保品質一致性。鼓勵資訊科學、教育學、人文社會等領域共創AI教育應用模型，使其更貼近高教現場需求與價值觀。並將AI納入高教改革政策主軸，結合數位學習、教師發展、學生評量、國際評鑑等項目，形成整合性規劃與預算支持。

結語：共構AI時代的高教品保評鑑新典範

生成式AI的興起，為高等教育品質保證注入新動能，也挑戰既有制度與角色的邊界。在此轉型關鍵時刻，科技應為人文服務，AI應由教育理念引導。未來的高教品質保障，不應是科技主導人、亦非人抗拒科技，利用AI工具並不是偷懶也不是放棄人所擁有的智慧與判斷。而是透過人機協作、價值引導、制度創新，共同建立一個具備透明性、公平性與前瞻性的教育品質保障評鑑體系。在這條道路上，生成式AI不是終點，而是一種工具與契機，端看我們如何善用其力，落實教育的本質與使命。未來，我們應持續以開放、合作、倫理與專業的態度，迎向AI與高教融合的新世紀。