

人工智慧在主要科學教育期刊之相關研究： 文獻回顧與展望

張家榮¹ 楊曉菁^{2,*} 李良一³

¹元智大學 資訊傳播學系

²國立彰化師範大學 國文學系

³國立臺灣師範大學 學習科學學士學位學程

摘要

本研究旨在透過系統性文獻分析，回顧過去三個主要科學教育期刊 *International Journal of Science Education (IJSE)*、*Journal of Research in Science Teaching (JRST)* 及 *Science Education (SE)* 在人工智慧相關研究的發展趨勢，分析主要分成整體發展趨勢、實證研究發展趨勢、及非實證研究所探討的議題。本研究經有系統的篩選後針對25篇論文分析，結果顯示這些論文的時間分布大致反映出AI的浪潮。而25篇中的15篇有實證研究，實證研究發展趨勢主要依人工智慧應用類型、使用的人工智慧技術或工具、研究設計、研究方法、教育程度面向分析及報告，結果顯示人工智慧在學習者側寫及預測、評量及評估兩個應用類型最多，自然語言處理及決策樹為最常使用的人工智慧技術，實驗設計為最常使用的研究設計，傳統測驗為最常使用的研究方法，而國中為最常被調查的教育程度。而25篇中的10篇沒有實證研究，這10篇文章主要依以下三個主題討論及報告：人工智慧對科學教育評量的潛在挑戰與省思、人工智慧對科學教學與學習的潛在挑戰與省思、人工智慧對科學教育學術研究和期刊的潛在挑戰與省思。經由實證研究及非實證研究論文的分析及討論，本研究最後亦針對未來研究與科學教育方面提供幾點參考建議。

關鍵詞：人工智慧、文獻回顧、科學教育

壹、前言

人工智慧(artificial intelligence, AI)早在1956年就已形成一個新的研究領域，它的發展經過了幾次大的起伏，由早期1950~1960年代認為AI是搜尋與探索的浪潮，到1970年代認為AI就是輸入知識與推論的專家系統浪

潮，再到2000年後極大資料庫及機器學習的浪潮，這幾次浪潮都一步步推進AI的研究方向及應用範圍(Russell & Norvig, 2021)。而在2012年後深度學習演算法的改良、GPU (Graphics Processing Unit, 又譯圖形處理器) 高速計算的發展、網際網路收集的大數據資

*通訊作者：楊曉菁，jason0209@cc.ncue.edu.tw；ORCID：009-0002-5239-0322

投稿：2024/4/29，修訂：2024/7/29，接受：2024/7/30，線上出版：2024/11/29

料，這些的加成都讓AI能夠提供快速的計算及即時及準確的回應，也因此再次讓AI的技術擴大應用在我們的生活及工作，如自動駕駛、自動工廠及ChatGPT (Chat Generative Pre-trained Transformer)，而此次的浪潮還沒終止，而是繼續的在影響我們的社會及生活 (“Artificial intelligence,” 2024)。

在教育上AI也有很多的應用，Zawacki-Richter等(2019)分析了2007 ~ 2018年AI在高等教育的研究論文，將AI在教育的應用分成四個類別：一、學習者側寫及預測(profiling and prediction)，如建立學習者模型(learner model)及預測學生休學的預測；二、智慧型教學系統(intelligent tutoring systems)，如推薦學習者學習教材及診斷學習者的知識強弱處；三、評量及評估(assessment and evaluation)，如自動評分及評量學生的學習投入度；四、適性化系統及個人化(adaptive systems and personalization)，如提供個人化的學習回饋。這些應用都是使用不同的AI技術或工具來達成智慧化的教學目的。而在最近的一次AI浪潮後，新的AI工具被發展出來，研究除了傳統AI在教育的應用外，近幾年也漸漸將新的AI技術及工具應用在教學上如ChatGPT及自動翻譯等，並也開始討論AI對社會經濟的衝擊，並提出其對未來教學的看法，也因此近幾年AI在教育的相關研究也變得越來越多。為了促進AI在教育研究的發展，對於這些研究也應有系統的整理，來瞭解過去研究的狀況，並提出AI在教育應該注意的問題及未來可能的研究方向。

對於過去AI在教育上的研究，已有幾個文獻分析論文對其分析及探討，如Zawacki-Richter等(2019)依據年分、期刊、國家、作者單位、AI在教育的挑戰及危機、AI的應用，

分析了2007 ~ 2018年AI在高等教育的146篇研究論文。Jia等(2024)依據年分、主題、期刊、國家、被評量的學習成果、對教育的影響，分析了2013 ~ 2023年AI在科學教育的76篇研究論文。Crompton等(2024)依據學科、教育程度、研究目的、年分、國家、目標使用者、AI的應用、AI在教育的挑戰，分析了2011 ~ 2021年AI在K-12的146篇研究論文。Zhou等(2024)依據來源期刊、國家、年分、教育程度、AI的應用、使用的AI演算法、研究方法、國家、學習目的、年分、目標使用者、AI在教育的應用、AI在教育的挑戰及隱含，分析了2010 ~ 2023年AI在體育教育的130篇實證研究論文。Hwang等(2024)依據來源期刊、國家、AI的應用、使用的AI演算法、研究方法、學習活動、研究對象、被評量的學習成果，分析了1996 ~ 2020年AI在護理教育的112篇實證研究論文。

過去文獻分析論文專注在不同的教育程度及不同學科來收集及分析過去文獻，但我們只有發現Jia等(2024)是分析AI在科學教育的論文，因此應有更多的文獻分析論文從不同角度瞭解AI在科學教育的研究狀況及趨勢，這主要有以下幾個原因，首先，Jia等只分析實證研究論文，雖然分析實證性研究，能讓我們瞭解AI在科學教育實踐上的影響，然而，隨著AI技術的迅速發展，衍生出許多新型態的產物與應用，在這個新的時空背景下，AI在科學教育扮演哪些應用角色？在應用方面可能面臨哪些挑戰與限制？AI整合至科學教育環境中，對教育者與學習者角色可能有哪些影響尚不清楚，其可能需要考量因素與面臨的挑戰也不明確。而一些非實證論文如立場論文(position paper)及理論論文(theoretical paper)會對某些議題提出特定的意

見、立場及理論框架(T.-J. Lin et al., 2024)，因此藉由非實證性研究分析，一方面可以彌補過去與現在關於AI在科學教育議題上的研究斷層，另一方面，這也有助於引導未來AI在科學教育中應該關注的焦點與努力的方向。第二，Jia等收集的論文是由Web of Science (WoS)及Scopus的2013年6月至2023年6月間AI在科學教育的論文，這些論文主要來源期刊如下*International Journal of Social Robotics* (5篇)、*ETR&D-Educational Technology Research and Development* (4篇)、*International Journal of Technology and Design Education* (4篇)、*Education and Information Technologies* (4篇)、*Frontiers in Psychology* (3篇)、*British Journal of Educational Technology* (3篇)、*Journal of Science Education and Technology* (3篇)及*IEEE Transactions on Learning Technologies* (3篇)，這些期刊主要為教育科技相關期刊，因此論文的內容可能較偏向教育科技研究者角度，而非科學教育研究者的角度。最後Jia等論文只收集到2023年6月，但AI論文在2023~2024年間有大量的發表，因此需對2023年6月後的論文進行分析。而本研究的論文收集到2024年3月，增加了接近一年的時間論文發表，且本研究除了分析實證研究論文也分析非實證研究論文。

此外本研究只從三本科學教育的期刊收集而來，包括：*International Journal of Science Education* (IJSE)、*Journal of Research in Science Teaching* (JRST)及*Science Education* (SE)，會只收集及分析此三期刊的主要原因如下，第一是這三本期刊是歷史較悠久的科學教育期刊，如JRST是在1963年創刊、SE是在1916年創刊、IJSE是在1979年創刊。第二是這三本期刊都是科學教育中已長期被收

錄在Social Science Citation Index (SSCI)的期刊(如JRST是在1982年被收錄在SSCI、SE是在1984年、IJSE是在1987年)，因此文章數較多(如至2023年在WoS中共收錄2,412篇JRST的論文、1,945篇SE的論文、3,500篇IJSE的論文)。第三是本研究會分析實證及非實證論文，且並未限制時間範圍，若將所有文章收集可能過多，且內容品質可能參差不齊，而本研究所指定的三本期刊是科學教育中有較高影響力的期刊，如2023年的JCR在“Educational & Educational Research”類別中的756個期刊中，JRST排名48、SE排名74、IJSE排名169，因此所收集及分析的文章皆是科學教育高品質的文章，所分析的結果也應具有很大的代表性。最後是過去也有部分關於科學教育系統性回顧的研究，也是只分析這三本期刊(M.-H. Lee et al., 2009; T.-C. Lin et al., 2014; T.-J. Lin et al., 2019, 2024)，這些系統性回顧文章，針對這三本期刊，分析不同年分區段的所有文章，應能代表科學教育研究在不同時間區段的一般研究狀況及趨勢。而本研究則針對這三本期刊，分析收錄在WoS關於AI的所有文章，因此應對AI在科學教育研究的狀況及趨勢有一定代表性。

為了瞭解AI在此三個被選擇的期刊中，過去研究的趨勢及範圍，AI在科學教育需要考量因素與面臨的挑戰、及未來可能的研究方向，本研究主要回答以下幾個問題：

- 一、整體而言，AI在此三本被選擇的期刊中，依據時間發展的趨勢為何？
- 二、整體而言，AI在此三本被選擇的期刊中，實證研究所著重之AI應用類型、使用的AI技術或工具、研究設計、研究方法、教育程度為何？

三、整體而言，AI在此三本被選擇的期刊中，非實證研究文章主要探討的議題為何？

貳、研究方法

本研究相關文獻資料收錄(data inclusion)程序採PRISMA準則。PRISMA全名為Preferred Reporting Items for Systematic Reviews and Meta-Analyses，係透過系統化程序，包括：一、文獻檢索與識別 (identification)；二、文獻篩選 (screening)；三、符合標準 (eligibility)；四、最後收錄文獻(included)等步驟，提供穩健的資料來源進行回顧研究 (Moher et al., 2009)。流程如圖1，而各步驟的相關內容說明如下：

首先到WoS搜尋，在標題(title)、摘要(abstract)、關鍵字(keyword)欄位，使用以下關鍵字搜尋，“artificial intelligence”、

“AI”、“AIED”、“machine learning”、“intelligent tutoring system”、“expert system”、“recommended system”、“recommendation system”、“feedback system”、“personalized learning”、“adaptive learning”、“prediction system”、“student model”、“learner model”、“data mining”、“learning analytics”、“prediction model”、“automated evaluation”、“automated assessment”、“robot”、“virtual agent”、“algorithm”、“machine intelligence”、“intelligent support”、“intelligent system”、“deep learning”、“AI education”，限制期刊為IJSE、JRST及SE。會使用這些關鍵字主要有兩個原因，第一是在最初的搜尋中使用通用關鍵字“artificial intelligence”及“AI”進行文獻搜尋，但所得到的結果過少只有11篇，因此參考過去AI科學教育系統性文獻回顧研究中使用的關鍵字，因此，本研究使用如上之

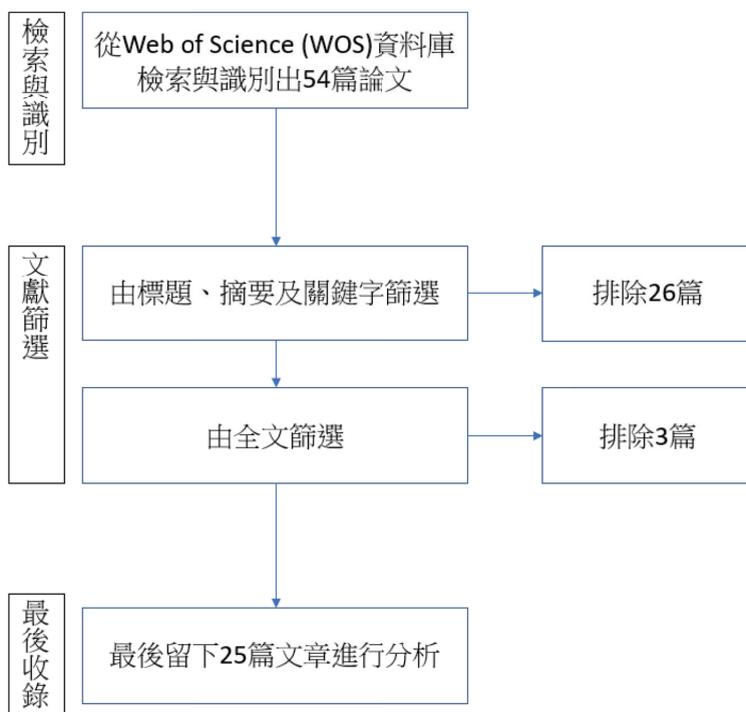


圖1：文章選擇流程圖

關鍵字，這些關鍵字與Jia等(2024)研究中所使用的搜尋關鍵字是相同的。

使用以上關鍵字搜尋出54篇論文，本研究將54篇論文進行兩階段篩選，第一階段篩選先由兩位研究者(第一及第二作者)閱讀54篇文章的標題、摘要及關鍵字，獨立判定該文章是否與AI相關，兩位研究者都需閱讀54篇文章的標題、摘要及關鍵字，對一篇文章，若兩位研究者皆判斷為與AI不相關的文章，則將此篇文章排除(如在摘要中有出現“deep learning”此關鍵字，但此處的“deep learning”不是AI的“deep learning”)，若只有一位研究者判斷為與AI不相關的文章，則仍保留此篇文章，最後在此階段共排除26篇文章，留下28篇文章。第二階段則由兩位研究者(第一及第二作者)閱讀28篇文章的全文，獨立判定該文章是否為實證研究的論文，評分者信度為88.9%。對於兩位研究者分類不一致的部分，兩位研究者進行討論來達成一致性，最後在此階段另外發現3篇與AI不相關的文章，將其排除，最後留下25篇文章，而最後收錄的25篇文章中，其中15篇有實證研究，10篇沒有實證研究。

實證文章依據以下幾個面向分析，包括：

- 一、AI應用類型；
- 二、使用的AI技術或工具；
- 三、研究設計；
- 四、研究方法；
- 五、教育程度。

以下將詳細說明這些面向及每一面向包含的類別。

- 一、AI應用類型：AI應用類型是參考Zawacki-Richter等(2019)所提出的分類法，包含四個類別：

- (一)學習者側寫及預測(profiling and prediction)；
- (二)智慧型教學系統(intelligent tutoring systems)；
- (三)評量及評估(assessment and evaluation)；
- (四)適性化系統及個人化(adaptive systems and personalization)。

會參考Zawacki-Richter等所提出的分類法主要有兩個原因，第一是Zawacki-Richter等的論文中有對每個類別有非常詳細的說明，這可幫助我們產生較一致的分類結果，第二是有幾個文獻分析的研究也使用此分類法，如Hwang等(2024)及Bond等(2024)。除了這四個類別外在內容編碼的過程中兩位編碼者也增加了兩個編碼：

- (五)研究資料分析(research data analysis)
- (六)對AI的觀點／感受／經驗(view/perception/experience to AI)。

以下為各類別的說明：

- (一)學習者側寫及預測：此類別的應用為建立學習者模型(learner model)作為預測使用，如學生休學的預測。
- (二)智慧型教學系統：此類別主要是應用AI來支援教學，如推薦學習者學習教材及路徑。
- (三)評量及評估：此類別主要是應用AI來支援學習評量，如自動評量系統。
- (四)適性化系統及個人化：此類別主要是應用AI來提供個人化及適性化的教學，如提供個人化的學習回饋。
- (五)研究資料分析：此部分研究主要使用AI來分析科學教育的研究資料，如大型資料庫(PISA資料庫)的資料分析。
- (六)對AI的觀點／感受／經驗：此部分研究

主要調查學生和老師對AI的觀點、感受及過去經驗。

二、使用的AI技術或工具：使用的AI技術或工具的類別產生經過兩個步驟，首先先將文章中所使用到的AI技術及工具記錄下來，之後由編碼者歸納這些文章中所使用的技術及工具主要為五種類別，包括以下幾個類別：

- (一)自然語言處理(natural language process)
- (二)群聚分析(clustering analysis)
- (三)決策樹(i.e., random forest and C4.5)
- (四)類神經網路(artificial neural network)，此類別是以類神經網路為基礎的演算法，如Convolution Neural Network(CNN)
- (五)多元回歸分析(multiple regression analysis)
- (六)其他，除了以上五種類別外，剩下的都歸類到此類別，此類別的技術及工具都只有一篇論文有使用，如支援向量機(support vector machines)。

三、研究設計：研究設計則依據Bryman (2004)對研究設計的定義及分類來進行編碼，分為以下幾個類別：

- (一)實驗設計(experimental design)
- (二)橫斷面研究設計(cross-sectional design)
- (三)縱貫研究設計(longitudinal design)
- (四)個案研究設計(case study design)

四、研究方法：研究方法也是依據Bryman (2004)對研究方法的定義及分類來進行編碼，分為以下幾個類別：

- (一)傳統測驗
- (二)問卷
- (三)訪談

(四)觀察

(五)文件分析

五、教育程度：

- (一)幼稚園至國小(k-elementary)
- (二)國中(middle)
- (三)高中(high)
- (四)大學(college)
- (五)其他

在此編碼架構中，「三」採單一編碼(singular coding)，「一」、「二」、「四」及「五」則採用多重編碼(multiple coding)。

參、研究結果

一、整體趨勢

本研究最後收錄的25篇文章中，如圖2所示來源期刊為7篇來自IJSE(28%)、15篇來自JRST(60%)、3篇來自SE(12%)，有超過一半以上的論文來自JRST，因此可看出JRST收

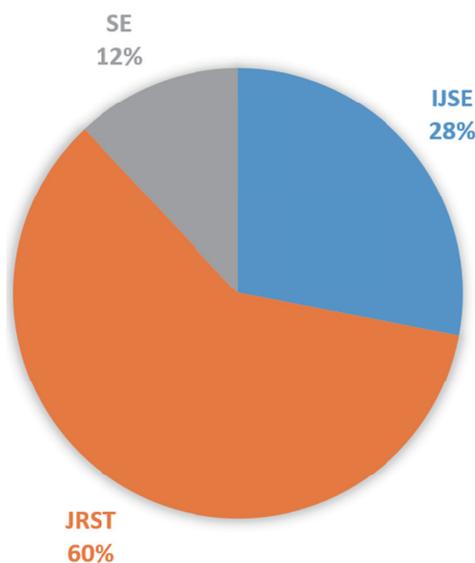


圖2：25篇論文的來源期刊分布

錄較多AI相關文章，而SE較少。這25篇論文中，15篇是實證研究(60%)、10篇是非實證研究(40%)。由時間順序來看這些論文(圖3)，1987年1篇非實證(Good, 1987)、1994年1篇實證(Andaloro et al., 1994)、2008年1篇實證(Liu & Ruiz, 2008)、2012年1篇實證(Yeh et al., 2012)、2015年1篇實證(Vitale et al., 2015)、2019年1篇實證(H.-S. Lee et al., 2019)、2020年1篇非實證(Zhai et al., 2020)、2021年4篇(2篇實證[Song & Wang, 2021; Vyas et al., 2021]、2篇非實證[Cheuk, 2021; Odden et al., 2021])、2022年2篇(1篇實證[Zhai et al., 2022]、1篇非實證[Wu & Tsai, 2022])、2023年6篇(3篇實證[Karahan, 2023; Körpeoğlu & Yıldız, 2024; Martin et al., 2024]、3篇非實證[Kubsch et al., 2023; Li et al., 2023; Zhai & Nehm, 2023])、2024年6篇(4篇實證[Ariely et al., 2024; Tan et al., 2024; Wang et al., 2024]、2篇非實證[Erduran & Levrini, 2024; Sadler et al., 2024])。

二、實證研究分析結果

此部分呈現15篇來自IJSE、JRST、SE關於AI的實證研究分析結果，結果依據「AI應用類型」、「使用的AI技術或工具」、「研究設計」、「研究方法」與「教育程度的整體趨勢」，其結果如表1所示。首先，從AI應用類型面相來看，結果顯示學習者側寫及預測(6篇，40.00%)與評量及評估(6篇，40.00%)是最多AI應用的類型，學習者側寫及預測的6個研究中，5篇是關於預測及1篇是建立學習者模型，在5篇關於預測的論文中，2篇是預測學習成效、1篇是預測STEM (Science, Technology, Engineering & Math)態度、1篇預測跨領域能力(interdisciplinary competence)、1篇預測學業復原力(academic resilience)，研究在預測的應用中，預測的目標都不相同。評量及評估的6個研究中，皆為提出自動評量的系統及方法，並評量其效果，但6個研究所評量的目標物也所不同，4篇為自動評量學生對

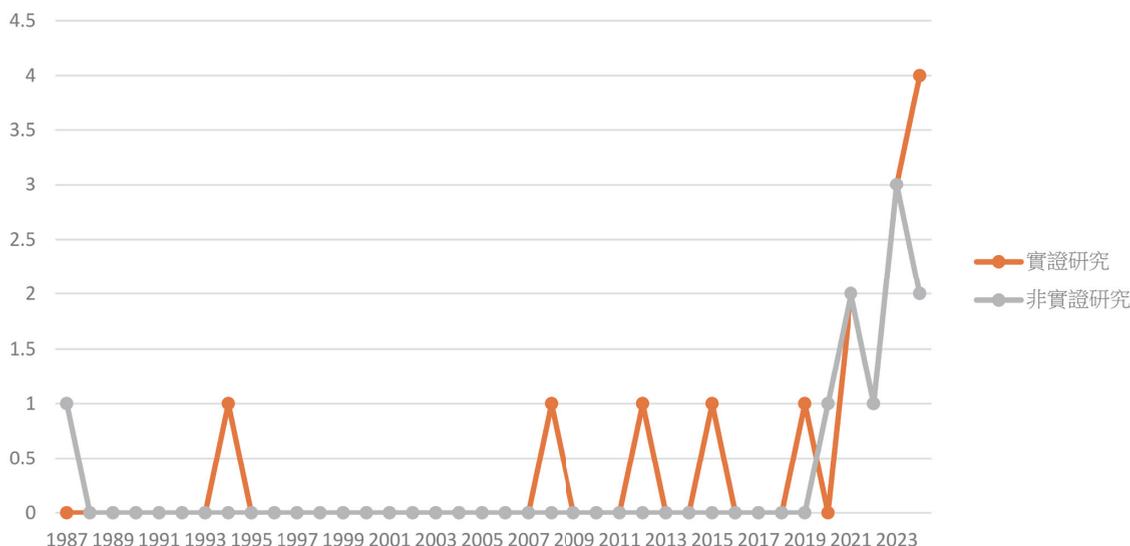


圖3：每年實證及非實證論文的個數

表1：實證研究整體分布情形

面向	類別	數量	百分比
一、AI應用類型 ^a	(一) 學習者側寫及預測	6	40.00
	(二) 智慧型教學系統	1	6.67
	(三) 評量及評估	6	40.00
	(四) 適性化系統及個人化	1	6.67
	(五) 研究資料分析	2	13.33
	(六) 對AI的觀點／感受／經驗	1	6.67
二、使用的AI技術或工具 ^a	(一) 自然語言處理	4	26.67
	(二) 群聚分析	3	20.00
	(三) 決策樹	4	26.67
	(四) 類神經網路	3	20.00
	(五) 多元回歸分析	2	13.33
	(六) 其他	5	33.33
三、研究設計	(一) 實驗設計	7	46.67
	(二) 橫斷面研究設計	4	26.67
	(三) 縱貫研究設計	1	6.67
	(四) 個案研究設計	3	20.00
四、研究方法 ^a	(一) 傳統測驗	8	53.33
	(二) 問卷	3	20.00
	(三) 訪談	3	20.00
	(四) 觀察	3	20.00
	(五) 文件分析	4	26.67
五、教育程度 ^a	(一) 幼稚園至國小	3	20.00
	(二) 國中	9	60.00
	(三) 高中	4	26.67
	(四) 大學	4	26.67
	(五) 其他	1	6.67

註：^a為複選題。

科學論證或解釋所寫的內容、2篇評量學生化
化的科學模型。

有2篇(13.33%)文章則使用AI技術進行
研究資料的分析，其中一篇分析WoS收集的
171篇與科學探究(scientific inquiry)相關的文章摘要，以瞭解主要的研究方向，另外一篇
則分析2015 PISA資料庫，以瞭解哪些因素
影響學生科學相關的職業期望。相較之下，
智慧型教學系統及適性化系統及個人化的關

注相對較少，智慧型教學系統只有1篇論文
(6.67%)，此篇論文是介紹一個能自動提供學
習回饋的系統，而適性化系統及個人化也只
有1篇論文(6.67%)，此篇論文是介紹及評量
一個能提供個人化學習回饋的系統。值得注
意的是有1篇(6.67%)實證研究論文並非AI應
用在教學及研究，而是發掘pre-service老師
(師資生)對AI的觀點。

在使用的AI技術或工具面向上，有使

用自然語言處理(4篇, 26.67%)、群聚分析(3篇, 20.00%)、決策樹(4篇, 26.67%)、類神經網路(3篇, 20.00%)、多元回歸分析(2篇, 13.33%)，此外有5篇文章中有使用其他的技術或工具，如支援向量機(support vector machines)、模糊邏輯(fuzzy logic)、VELO等，雖然研究有使用不同的AI技術或工具，但有約一半的研究(7篇, 46.67%)使用不只一種AI技術或工具在其研究中。在研究設計面向，主要使用實驗設計(7篇, 46.67%)，其次是橫斷面研究設計(4篇, 26.67%)、個案研究設計(3篇, 20.00%)及縱貫研究設計(1篇, 6.67%)，實驗設計會較多可能是大部分研究需評量AI應用或AI工具的效能，因此研究設計是為了評估效能。在研究方法面向上，結果顯示一半以上的研究使用傳統考試的方式(8篇, 53.33%)進行研究資料的收集，其次使用文件分析(4篇, 26.67%)，接著是問卷(3篇, 20.00%)、訪談(3篇, 20.00%)和觀察(3篇, 20.00%)。最後，在教育程度面相來看，這些相關研究實施橫跨不同的教育程度，結果顯示研究最廣泛的教育程度是國中(9篇, 60.00%)，其次為高中(4篇, 26.67%)、大學(4篇, 26.67%)及幼稚園至國小(3篇, 20.00%)，最後是其它(1篇, 6.67%)。

三、非實證研究分析結果

此部分呈現10篇來自IJSE、JRST、SE關於AI的非實證研究內容分析結果，結果顯示在AI在科學教育的應用上，有7篇(70.00%)有探討評量與評估，其次有4篇(40.00%)有探討適性化系統和個人化、有3篇(30.00%)有探討學習者側寫及預測，有2篇(20.00%)有探討研究資料分析、有1篇(10.00%)有探討智慧型教學系統。雖然這些論文有探討AI在科學教育的不同應用，但這些文章主要圍繞在以下3個主題討論：(一) AI對科學教育評量的潛在挑

戰與省思、(二) AI對科學教學與學習的潛在挑戰與省思、(三) AI對科學教育學術研究和期刊的潛在挑戰與省思。

(一) AI對科學教育評量的潛在挑戰與省思

在10篇非實證性研究中有7篇(70.00%)論文有探討AI對科學評量的影響，其主要使用機器學習技術自動化評量學生分數。Zhai等(2020)的文章中比較使用機器學習和一般科學評量之間相似與差異，並指出使用機器學習方式可以其藉由多種證據推論學生的認知、情感和感覺狀態，並減少教師人工評分的負擔。而Zhai與Nehm (2023)的研究也指出AI對形成性評量可以提供有價值的回饋給教師進而制訂教學計畫以及決策。除了正面的影響之外，這些研究中也指出使用AI應用於自動評分上的潛在問題與挑戰。例如Zhai等說明AI衍生的問題：

However, most of the algorithms (68%) are content specific, which means they could only be applied to the same item and are not generalizable for other contexts. (p. 1446)

Among the studies, only eight did not require humans to label the data to build the algorithm, which means that most studies used a supervised ML approach (39 articles) or semi-supervised ML approach (one article), which requires significant human effort. (p. 1446)

這些問題指出目前機器學使用的演算法因內容具有特定情境，因此無法適用於一般的情境。而且需要耗費大量的人力進行資料標記進而訓練模型，因此增加開發成本與時間，並且建立的評量模型具有侷限性問題。

同樣對於演算法問題，Cheuk (2021)指出機器學習可能因為使用的演算法而產生不平等的結構，因為演算法是人類設計而產生，無法避免傳遞設計者有意識或無意識的信念和偏見，在預先已存在的偏見模型中可能清除了一些不被欣賞的學生族群，導致偏見持續存在。而其他研究則針對人工標記過程進一步不同的討論議題。不同於訓練模型的資料集來源問題，Kubsch等(2023)對於AI偏見則聚焦在人為驅動所導致，其介紹兩種不同機器學習分析資料過程中人類扮演的角色而產生的偏見。更具體來說，一種方式是人類事先制定並檢驗評分標準，此標準作為機器學習複製人類評分的過程；而另一種方式則是使用非監督式機器學習進行詞彙分析找出模式，也就是自然語言處理，後續則交由人類透過深度閱讀和內容進行模式修改。

Kubsch等(2023)提出如果人類使用機器學習進行自動化評分但卻不清楚人類在計算分析過程中扮演的角色，將衍生倫理與透明化問題，以及方法論整合的本質問題。前者是指若不清楚人類在分析過程中的角色，或是人類和演算法之間關係，將會無法分辨偏見的來源；後者是不清楚使用AI自動評分的目的是為了複製人類評分的行為，亦或是為了增強人類建立新的洞察表現。Zhai與Nehm (2023)進一步指出目前常見的AI偏見(bias)具有三種類型，分別是偏差(deviation)、系統性錯誤(systematic)和傾向(tendency)。因為AI存在偏見問題，Li等(2023)認為由於不瞭解自動評分的審查過程，偏見問題可能造成學習上的不公平，另一方面則是因為訓練模型的數據集可能主要來自主流學生族群，可能導致減少文化相關的教學，並忽略對少數學生族群的文化和語言。

為了解決AI偏見問題，相關研究從各種

觀點給予建議，Zhai與Nehm (2023)作者建議AI不應作為一個問題解決的工具，而是應該視為促進者的角色與合作對象。這樣的觀點與Li等(2023)一致，其建議AI研究者可以探索人機合作的方式，讓教師和自動化系統共同合作(human-in-the-loop approaches)，結合教師的優勢和AI的效率來提升評量的過程。而Cheuk (2021)則建議讓批判學者與機器學習評估設計者以交叉方式共同合作，並組織一個演算法正義聯盟呈現演算法的公平與透明，藉由批判交叉性框架可以將少數學生族群作為機器學習評量設計的中心，以避免少數學生族群被排除的偏見問題。而Zhai與Nehm則提出不同的觀點，其認為沒有單一評量工具可以全面性捕捉學生的想法和背景，即使老師也難以達到，其表示AI評分的偏見不等於不公平教育，如果AI能夠產生可靠並即時的評估資訊，反而有助於老師幫助有困難的學生。Kubsch等(2023)期望科學教育研究能更具創造性地探索機器學習在科學教育研究中的角色，而不僅僅局限於自動評分的應用。

(二)AI對科學教育與學習的潛在挑戰與省思

儘管許多研究應用AI於自動評量並探討AI使用的技術與挑戰，有另外一部分研究學者則是關注AI對於教學與學習的影響，其討論範疇包含教師對AI的理解與使用能力、扮演角色的定位、教學目標以及責任。這些文獻透過不同層面的探討可以幫助科學教育者以及研究者對AI於未來科學教育有更深入的瞭解。早期Good (1987)探討機器學習可以作為專家系統診斷學生學習表現或迷思概念，以及模仿教師的診斷行為，然而因為自然語言處理、模仿學生學習和教師建模三個因素導致在科學教育中專家系統的發展緩慢。其中影響科學學習有很多且複雜的因素，Good進一步提出不管指導者為機器或是教師，如

何使用準確的診斷幫助學生學習取決於下一個學習目標。換句話說，AI的應用需要有合適的教學策略才能提供有效的學習。而Zhai等(2020)則認為AI不僅只是診斷學生認知表現而已，其提供教師相關資訊可作為決策的相關資訊。但Zhai等進一步認為目前機器學習在教育決策上仍存在一些挑戰，因為其提供的資訊或證據仍缺乏某些類型，例如認知、情緒和感官等資訊，這些相關資訊在教育決策上也是相當重要的決定因素。

除了應用AI學習教師的評分過程之外，Zhai等(2020)也建議AI可使用行為、社會和文化多種資訊提供提升更準確的教學決策，以支持科學學習的適性化和個人化學習需求。同樣地，Li等(2023)則建議自動評分可進一步來加深和個人化教師與學生互動的方式。對於AI支援個人化和適性化學習，Wu與Tsai(2022)提出AI目前針對個人化學習仍有限制與挑戰，其指出雖然AI技術可以自動評分或是偵測個體和網路知識信念，以及建模學習者和作品之間不同的關係，但目前多數研究在有限的形式和來源分類學生的知識信念，並且多數研究著重於文本寫作分析，較少研究針對不同形式的繪圖或是口語的模式辨識。因此，Wu與Tsai進一步基於增強學習提出人機共生學習分析架構以提升AI提供更多的訊息輔助教學設計。這樣的架構主要協調社會網路、智慧型教室、智慧型代理人和儀表板之間的訊息處理，藉由捕捉學生多模資料建立即時自動反饋和鷹架以支援個人化學習。

雖然上述這些研究表示AI支援個人化學習有很大的潛力，然而一些研究學者進一步提出如何有效地整合AI於教學的議題，例如Zhai與Nehm(2023)提出如何訓練教師使用並瞭解AI的優點與限制，以及AI如何廣泛

地適用於各種教學和學習情境。其建議未來研究應該提供教師對於AI影響更全面性的理解，建立實踐指南以促進教師更佳地制定AI在課堂中使用。Wu與Tsai(2022)也關注教師訓練問題，其提出教師專業發展是否應該包含理解AI以及教師如何與智慧代理人一起合作。除此之外，Wu與Tsai進一步提出如果AI能夠評量學生的作答，並且評量學生的知識思考後提供立即性回饋，在教學過程中教師要扮演什麼角色？對於學生而言，科學教育的整體目標是培養學生的批判性思考以及合作問題解決能力，他們基於適應化知識信念以知識是不確定的特徵，科學學習應該是藉由實驗或依賴任務情境的多個來源來證明進行驗證。若學生在AI的介入下提供及時鷹架和回饋，學生僅是接收提供的回饋，是否會造成學生發展較少的知識信念呢？學生不需要請求幫助或是承擔其學習責任，學生是否會變成被動學習者呢？因此，許多研究建議未來科學教育需要聚焦在AI在課堂、教師教育中的有效整合，以及開發能夠適應多樣化教學環境的AI系統(Wu & Tsai; Zhai & Nehm, 2023)。

(三)AI對科學教育學術研究和期刊的潛在挑戰與省思

除了探討AI在教育應用之外，少部分研究學者開始探索使用AI幫助研究者分析與組織龐大的文獻資料。例如在Odden等(2021)研究中採用機器學習分析和預測某特定期刊的主題，相較傳統以研究為焦點的文獻方式，其結合自然語言處理以及分類技術，可以全面性偵測和分析近100年期間科學教育和學習的演進與變化。更具體來說，他們使用自然語言處理領域中的自動化機器學習演算法針對科學教育研究語料庫進行文獻綜術與分類，分析結果顯示過去研究主要有21種

主題，他們進一步群聚分成科學內容主題、以教學為中心和以學生為中心三大主題，藉由時間分析可以瞭解每個大主題的發展與趨勢。

雖然AI可以幫學術研究進行複雜且大量的數據分析與組織，但這樣的分析方式也有其優勢與限制，Odden等(2021)對於潛在問題也進一步提出說明，研究結果主要是使用特定數學模型或是演算法依據提供的研究資料進行分析，因此分析結果受限於對目前收集的資料集的詮釋。基於此概念，這樣的分析結果可展現研究主題的趨勢發展以及關注程度，但無法提供質性上更詳細的解釋。Odden等解釋採用特定方式分析僅能呈現某主題在數據量化上的變化程度，但是無法說明其重要程度。他們認為雖然使用機器學習分析的結果具有說服力，可以強化我們對於相關文獻的理解，但有些個別主題或文章的價值與優點可能無法藉由數據的變化展現其重要性。因此，他們不建議使用機器學習取代人類解釋的需求。

除此之外，Sadler等(2024)則關心AI對學術研究與出版社帶來的影響，其提出兩個基本主張，生成式AI是改寫或是組織基於現有資料，並非創造新的知識，這對於學術研究目的來說有所衝突。其次，生成式AI的處理流程非透明化，因此不清楚AI生成的資料來源、使用什麼方式分析資料、為何有些想法被強調但有些並沒有，因此面對生成式AI的潛在問題，他們提出學術研究面對AI的一些指導準則：“In terms of guidelines for the journal regarding AI, transparency is our key principle.” (p. 741), “Whereas the use of AI tools for the preparation of manuscripts should be clearly acknowledged, these tools cannot be included as coauthors in JRST.” (p. 741)

Sadler等(2024)針對審查議題也提出其看法，其認為AI目前的能力無法取代人工論文審稿過程，因為審查回饋是基於專業知識而非是AI生成，他們表示使用AI在學術研究或出版上仍有侷限性：

The editorial team sends manuscripts to reviewers to read and provide feedback based on their expertise, and we expect the feedback provided to be the product of the expert reviewers and not AI. We think that reviewing and making publication decisions on science education research manuscripts requires specialized knowledge and that current AI tools cannot complete these tasks well nor do they currently have the capacity to do so. (p. 742)

肆、討論

本研究收集三個主要科學教育期刊JRST、SE、IJSE中關於AI的文章，其包含實證性與非實證性研究，並回答三個研究問題。對於第一個研究問題，我們發現這些論文的時間發展趨勢大致反映出AI的浪潮，本次分析收錄最早的論文發表在1987年的JRST，該論文為非實證論文，內容包括AI的定義、AI在科學教育可能的應用、AI在科學教育為來研究方向及挑戰。之後由1990~2019年間共有5篇論文，皆為實證研究論文，這5篇論文有3篇是評估AI的教學系統，如Suzuki等(1994)的教學診斷系統、Vitale等(2015)的自動評分系統、H.-S. Lee等(2019)的自動評分及適性化回饋系統，另外2篇則為應用機器學習技術分析科學教育研究資料，如Liu與Ruiz (2008)使用決策樹及M5方法，分析International Mathematics and Science Study

(TIMSS)及National Assessment of Educational Progress (NAEP)的大型資料，Yeh等(2012)使用群聚分析，分析171篇科學教育論文的摘要，這也可能反映了此時段AI新應用的停滯，只能應用現有的技術進行教育系統的加強及研究資料的分析。在2020年後AI論文快速的增加，這可能反映了2020後，由深度學習、GPU及大數據資料的加成，有了新的應用，讓AI再次崛起，也讓AI再次受到重視，這些論文也有很大比例的論文為非實證論文，探討AI對科學教育未來的挑戰及可能性。

對於研究問題二，實證性文章依照AI應用類型、使用的AI技術或工具、研究設計、研究方法、教育程度等維度進行內容分析，研究結果顯示AI在學習者側寫及預測(40.00%)、評量及評估(40.00%)兩個類型最多，這樣的結果與Zawacki-Richter等(2019)研究的結果相似，其研究指出多數教育研究的AI應用類型著重在學習者側寫與預測(38.67%)，其次為評量與評估(24.00%)。與Zawacki-Richter等比較，本研究發現在科學教育中對於評量及評估面向有提高的趨勢(24.00% → 40.00%)，而智慧型教學系統面向有下降趨勢(19.33% → 6.67%)，適性化系統及個人化面向也呈現下降趨勢(18.00% → 6.67%)。由於Zawacki-Richter等主要分析高等教育的論文，因此這些結果可能代表科學教育較關心AI如何分析並預測學生的科學學習表現及如何應用AI進行科學學習自動化評分，而這些預測及評量的標的物也可能與科學教育較相關，如評量及評估面向的6個研究所評量的目標物中，4篇為自動評量學生對科學論證或解釋所寫的內容、2篇評量學生化化的科學模型。

從AI技術面向來看，研究所使用的AI工

具或技術可能跟AI應用類型有關，因此本研究進一步分析AI應用與AI工具或技術的交叉分析，我們發現學習者側寫及預測主題多數採用決策樹以及多元回歸分析，其中決策樹技術主要使用分類規則判斷學生的學習狀態或是診斷迷思，而多元回歸分析則運用數學模型預測學習者的表現。而自然語言處理主要應用在評量及評估，這可能因為在評量及評估的應用中，所評量的內容大部分為學生對科學論證或解釋所寫的內容，這些分析需要用到自然語言處理技術，但其他應用類型在使用的AI工具或技術則沒有顯著的趨勢。

在研究設計及方法面向上，主要使用實驗設計，而研究方法則是傳統考試的方式較多，而文件分析次多，傳統考試的方式較多可能呼應大部分研究設計為實驗設計，而文件分析次多則反映在評量及評估的應用中，被評量的內容大部分為學生對科學論證或解釋所寫的內容。在教育程度面向上，相較於Zawacki-Richter等(2019)研究著重於高等教育有所不同，本研究結果發現AI橫跨不同的教育程度面，以國中教育程度為主要應用階段，其次為高中與大學、幼稚園至國小。這些結果反映各國教育政策對AI的重視，呈現越來越廣泛的趨勢，並且在學前與國小教育階段越來越受重視。

對於研究問題三，在非實證性研究結果顯示，AI在科學教育應用主題主要為評量與評估，其次為適性化系統和個人化、學習者側寫和預測。與實證性研究相較之下，兩者研究皆顯示對評量與評估有高度的關注，然而適性化系統與個人化主題則有明顯的落差。在評量與評估主題上，非實證性研究顯示潛在值得注意的議題，例如演算法的侷限性、評量不透明化、存在資料集與人為偏見等挑戰，少數學生族群資料未被分析，相關

研究也提出共同提出人機合作模式解決方案。在適性化系統和個人化主題方面，非實證性研究認為目前的AI挑戰包含缺乏多元證據資訊幫助教師決定，例如缺少行為、社會、文化、情緒和感官等資訊。另外一個挑戰則是缺乏多元資料的分析，目前AI多數集中在文本分析，對於多元表徵類型(例如：繪圖、口語)等資料尚未提出分析模式。因此，有些學者提出人機共生學習分析架構，期望透過捕捉更多學生的多模資料以建立自動反饋與鷹架提升教學設計。然而，這些潛在議題在實證性研究中尚未重視與驗證，也不清楚這些理論解決方案的實施與可能面對的問題。這些結果提供未來研究一個值得關注的方向，如何設計人類與AI之間的合作模式或機制，此議題需要更多的研究學者投入研究共同提出可行的解決方案。

此外，非實證性研究進一步從教師與學習者觀點提出AI介入的潛在問題與挑戰，例如：教師在AI輔助下扮演的角色、教師專業發展的訓練、教師如何與AI合作等議題。對於學生而言，AI即時鷹架或回饋是否會造成被動學習問題，學生被動接受資訊是否會影響其對科學學習的知識信念。然而在實證性研究中尚未探討這些潛在的議題，建議未來可以針對科學教育實證研究進行更詳細的分析，提出教學實驗的實證性結果，以釐清AI在教學實務上的挑戰與設計準則。

除了探討AI在教學與學習議題之外，實證與非實證性研究皆有一些文章關心研究資料分析主題，其使用AI技術分析大量的研究文獻，從中萃取主要的研究概念與主題，進而瞭解整體發展趨勢與各主題之間的轉移。雖然AI在大規模研究文獻可以幫助研究者分類主題、時間分布的發展趨勢、區別研究主題的差異性等優點，但這些研究也指出使

用AI進行數據分析的限制，結果僅是模型對資料的解釋而非客觀上的真實，無法呈現個別研究主題的價值與重要性。因此，研究者不建議使用AI取代人類解釋。而在非實證性研究中，僅有一篇文章談論生成式AI對期刊與出版社可能帶來的影響與挑戰，其以學術研究文章目的與審查倫理觀點，指出生成式AI的資料集來源不清楚、生成內容非創造新的知識、缺乏透明化的審查過程等問題。因此，研究者同樣不建議使用AI取代人工文章撰寫與文章審查。

伍、未來研究展望與建議

綜合上述，本研究針對科學教育三個主要期刊文章作為研究分析樣本，以AI在科學教育上的應用類型作為探討主軸，透過深入探討各面向的分布差異，除了能夠幫助後續研究瞭解AI在科學教育近年的發展主軸與脈絡之外，並藉由分析非實證研究歸納目前可能潛在的挑戰與仍待解決的問題。除此之外，基於本研究的發現可作為後續研究彌補理論與實證研究之間的落差，以實證研究作為相關文獻探討基礎，以非實證性研究提供可研究議題的洞見。因此，我們針對AI在科學教育提出以下幾點未來研究展望與建議：

一、擴展AI的應用類型：相關研究目前主要著重於評量與評估、學習者側寫與預測兩個類型，對於適性化系統與個人化主題相關研究較少。特別的是，較少的研究調查教學者或是學習者使用AI的觀點或經驗。未來研究可以探索更多元的潛在應用類型，例如適性化教學、智慧型教學系統、研究資料分析等等，並且調查學習者在這些類型的觀點與使用經驗，並探索學習者如何與AI進行互動學習。因此，了解學習者使用AI的經驗與

互動可以幫助未來研究者以及教育者更好地設計教學活動。

二、發展新的研究方法與技術：目前研究方法多數採用傳統測量方式，雖然有少數研究採用文本、觀察、問卷等方式評量學生狀態，一方面，對於多模態類型的測量仍缺乏足夠訓練資料，導致建立的模型無法廣泛地實施評量。另一方面，如何解決AI存在偏見問題也尚未提出實證的解決方案，特別是教師與演算法皆有可能產生偏見問題，而教師與AI如何有效的相互合作仍缺乏一套標準規範，以發展一套有效的評量模型解決目前現有的偏見問題。未來研究仍需要持續追蹤並發展適合的演算法以廣泛地綜合評量學習者表現，進而有效地整合AI於科學教育。

三、探討AI對教育者與學習者的角色與影響：未來研究可朝向深入分析與探索在AI介入下對教育者與學習者的影響。對於教師而言，是否具備足夠的專業能力使用AI技術於科學教育中，以及目前師資培訓與教師專業發展是否需要加入AI素養仍需要進一步探討。除此之外，現存文獻多數為直接應用AI分析學生學習表現，教師如何整合AI於科學教學活動仍未提出有效的教學設計。對於學生而言，教學者或是學習者使用AI的觀點、情感或是經驗，因為這些因素不僅可能會影響學習者科學學習的學習動機與學

習成效，而且可能會影響學習者對於科學知識與信念的學習方法。因此，在AI輔助下是否會影響學生知識信念、情感與學習行為的影響仍需要更多實證性研究驗證。除此之外，教師與AI的角色定位仍須要進一步討論，幫助教師與學生釐清兩者角色的價值與負責項目。

四、建立AI學術倫理與應用規範：AI在評量與學術研究主題上仍缺乏一套資料收集與標準規範，以及揭露不透明的AI分析過程，未來研究可設計一套人機共同設計合作機制或架構，將有助於研究者與教育者瞭解AI評量標準與限制，並改善AI存在偏見的問題。對於學術研究者而言，明確的使用指南與應用規範可以幫助研究者在文章撰寫過程中避免偏見和歧視性結果的產生，以提高研究透明度與可驗證性，以確保學術研究的準確性與客觀性。

這些展望和建議目的在於吸引更多對科學教育感興趣的研究學者共同參與，共同探索與應用AI技術於科學教育與學習。期望透過這些展望與建議，提供未來的研究者和教育者在科學教育領域中更深入地探索和應用AI技術，從而推動科學教育的創新與發展。

誌謝

本研究感謝國家科學及技術委員會經費支持(NSTC 113-2410-H-155-009-)，以及審查委員提供的寶貴意見，在此特致謝忱。

參考文獻

- Andaloro, G., Bellomonte, L., Lupo, L., & Sperandeo-Mineo, R. M. (1994). Construction and validation of a computer-based diagnostic module on average velocity. *Journal of Research in Science Teaching*, 31(1), 53-63. <https://doi.org/10.1002/tea.3660310106>

- Ariely, M., Nazaretsky, T., & Alexandron, G. (2024). Causal-mechanical explanations in biology: Applying automated assessment for personalized learning in the science classroom. *Journal of Research in Science Teaching*, 61(8), 1858-1889. <https://doi.org/10.1002/tea.21929>
- Artificial intelligence. (2024, April 26). In *Wikipedia*. <https://bit.ly/4fvJUIy>
- Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S. W., & Siemens, G. (2024). A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. *International Journal of Educational Technology in Higher Education*, 21, Article 4. <https://doi.org/10.1186/s41239-023-00436-z>
- Bryman, A. (2004). *Social research methods* (2nd ed.). Oxford University Press.
- Cheuk, T. (2021). Can AI be racist? Color-evasiveness in the application of machine learning to science assessments. *Science Education*, 105(5), 825-836. <https://doi.org/10.1002/sce.21671>
- Crompton, H., Jones, M. V., & Burke, D. (2024). Affordances and challenges of artificial intelligence in K-12 education: A systematic review. *Journal of Research on Technology in Education*, 56(3), 248-268. <https://doi.org/10.1080/15391523.2022.2121344>
- Erduran, S., & Levrini, O. (2024). The impact of artificial intelligence on scientific practices: An emergent area of research for science education. *International Journal of Science Education*, 46(18), 1982-1989. <https://doi.org/10.1080/09500693.2024.2306604>
- Good, R. (1987). Artificial intelligence and science education. *Journal of Research in Science Teaching*, 24(4), 325-342. <https://doi.org/10.1002/tea.3660240406>
- Hwang, G.-J., Tang, K.-Y., & Tu, Y.-F. (2024). How artificial intelligence (AI) supports nursing education: Profiling the roles, applications, and trends of AI in nursing education research (1993-2020). *Interactive Learning Environments*, 32(1), 373-392. <https://doi.org/10.1080/10494820.2022.2086579>
- Jia, F., Sun, D., & Looi, C.-k. (2024). Artificial intelligence in science education (2013-2023): research trends in ten years. *Journal of Science Education and Technology*, 33(1), 94-117. <https://doi.org/10.1007/s10956-023-10077-6>
- Karahan, E. (2023). Using video-elicitation focus group interviews to explore pre-service science teachers' views and reasoning on artificial intelligence. *International Journal of Science Education*, 45(15), 1283-1302. <https://doi.org/10.1080/09500693.2023.2200887>
- Kubsch, M., Krist, C., & Rosenberg, J. M. (2023). Distributing epistemic functions and tasks—A framework for augmenting human analytic power with machine learning in science education research. *Journal of Research in Science Teaching*, 60(2), 423-447. <https://doi.org/10.1002/tea.21803>
- Körpeoğlu, S. G., & Yıldız, S. G. (2024). Using artificial intelligence to predict students' STEM attitudes: An adaptive neural-network-based fuzzy logic model. *International Journal of*

- Science Education*, 46(10), 1001-1026. <https://doi.org/10.1080/09500693.2023.2269291>
- Lee, H.-S., Pallant, A., Pryputniewicz, S., Lord, T., Mulholland, M., & Liu, O. L. (2019). Automated text scoring and real-time adjustable feedback: Supporting revision of scientific arguments involving uncertainty. *Science Education*, 103(3), 590-622. <https://doi.org/10.1002/sce.21504>
- Lee, M.-H., Wu, Y.-T., & Tsai, C.-C. (2009). Research trends in science education from 2003 to 2007: A content analysis of publications in selected journals. *International Journal of Science Education*, 31(15), 1999-2020. <https://doi.org/10.1080/09500690802314876>
- Li, T., Reigh, E., He, P., & Miller, E. A. (2023). Can we and should we use artificial intelligence for formative assessment in science? *Journal of Research in Science Teaching*, 60(6), 1385-1389. <https://doi.org/10.1002/tea.21867>
- Lin, T.-C., Lin, T.-J., & Tsai, C.-C. (2014). Research trends in science education from 2008 to 2012: A systematic content analysis of publications in selected journals. *International Journal of Science Education*, 36(8), 1346-1372. <https://doi.org/10.1080/09500693.2013.864428>
- Lin, T.-J., Lin, T.-C., Potvin, P., & Tsai, C.-C. (2019). Research trends in science education from 2013 to 2017: A systematic content analysis of publications in selected journals. *International Journal of Science Education*, 41(3), 367-387. <https://doi.org/10.1080/09500693.2018.1550274>
- Lin, T.-J., Lin, T.-C., Potvin, P., & Tsai, C.-C. (2024). Research trends in science education from 2018 to 2022: A systematic content analysis of publications in selected journals. *International Journal of Science Education*. <https://doi.org/10.1080/09500693.2024.2340811>
- Liu, X., & Ruiz, M. E. (2008). Using data mining to predict K-12 students' performance on large-scale assessment items related to energy. *Journal of Research in Science Teaching*, 45(5), 554-573. <https://doi.org/10.1002/tea.20232>
- Martin, P. P., Kranz, D., Wulff, P., & Graulich, N. (2024). Exploring new depths: Applying machine learning for the analysis of student argumentation in chemistry. *Journal of Research in Science Teaching*, 61(8), 1757-1792. <https://doi.org/10.1002/tea.21903>
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of internal medicine*, 151(4), 264-269. <https://doi.org/10.7326/0003-4819-151-4-200908180-00135>
- Odden, T. O. B., Marin, A., & Rudolph, J. L. (2021). How has *Science Education* changed over the last 100 years? An analysis using natural language processing. *Science Education*, 105(4), 653-680. <https://doi.org/10.1002/sce.21623>
- Russell, S. J., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson.
- Sadler, T. D., Mensah, F. M., & Tam, J. (2024). Artificial intelligence and the *Journal of Research*

- in Science Teaching. Journal of Research in Science Teaching*, 61(4), 739-743. <https://doi.org/10.1002/tea.21933>
- Song, G., & Wang, Z. (2021). Factors influencing middle school students' interdisciplinary competence in science education. *Journal of Research in Science Teaching*, 58(7), 1041-1072. <https://doi.org/10.1002/tea.21692>
- Suzuki, T., Yasuda, T., Yokoi, S., & Toriwaki, J.-i. (1994). Construction of virtual town and driving simulation in it using graphics workstation. *Journal of the Institute of Television Engineers of Japan*, 48(10), 1318-1325. <https://doi.org/10.3169/itej1978.48.1318>
- Tan, L., Chen, F., & Wei, B. (2024). Examining key capitals contributing to students' science-related career expectations and their relationship patterns: A machine learning approach. *Journal of Research in Science Teaching*, 61(8), 1975-2010. <https://doi.org/10.1002/tea.21939>
- Vitale, J. M., Lai, K., & Linn, M. C. (2015). Taking advantage of automated assessment of student-constructed graphs in science. *Journal of Research in Science Teaching*, 52(10), 1426-1450. <https://doi.org/10.1002/tea.21241>
- Vyas, V. S., Kemp, B., & Reid, S. A. (2021). Zeroing in on the best early-course metrics to identify at-risk students in general chemistry: An adaptive learning pre-assessment vs. traditional diagnostic exam. *International Journal of Science Education*, 43(4), 552-569. <https://doi.org/10.1080/09500693.2021.1874071>
- Wang, F., King, R. B., Fu, L., Chai, C.-S., & Leung, S. O. (2024). Overcoming adversity: Exploring the key predictors of academic resilience in science. *International Journal of Science Education*, 46(4), 313-337. <https://doi.org/10.1080/09500693.2023.2231117>
- Wu, J.-Y., & Tsai, C.-C. (2022). Harnessing the power of promising technologies to transform science education: prospects and challenges to promote adaptive epistemic beliefs in science learning. *International Journal of Science Education*, 44(2), 346-353. <https://doi.org/10.1080/09500693.2022.2028927>
- Yeh, Y.-F., Jen, T.-H., & Hsu, Y.-S. (2012). Major strands in scientific inquiry through cluster analysis of research abstracts. *International Journal of Science Education*, 34(18), 2811-2842. <https://doi.org/10.1080/09500693.2012.663513>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? *International Journal of Educational Technology in Higher Education*, 16, Article 39. <https://doi.org/10.1186/s41239-019-0171-0>
- Zhai, X., Haudek, K. C., Shi, L., Nehm, R. H., & Urban-Lurain, M. (2020). From substitution to redefinition: A framework of machine learning-based science assessment. *Journal of Research in Science Teaching*, 57(9), 1430-1459. <https://doi.org/10.1002/tea.21658>

- Zhai, X., He, P., & Krajcik, J. (2022). Applying machine learning to automatically assess scientific models. *Journal of Research in Science Teaching*, 59(10), 1765-1794. <https://doi.org/10.1002/tea.21773>
- Zhai, X., & Nehm, R. H. (2023). AI and formative assessment: The train has left the station. *Journal of Research in Science Teaching*, 60(6), 1390-1398. <https://doi.org/10.1002/tea.21885>
- Zhou, T., Wu, X., Wang, Y., Wang, Y., & Zhang, S. (2024). Application of artificial intelligence in physical education: A systematic review. *Education and Information Technologies*, 29(7), 8203-8220. <https://doi.org/10.1007/s10639-023-12128-2>

Artificial Intelligence in Publications of Selected Journals: A Systematic Literature Review and Prospect

Chia-Jung Chang¹, Hsiao-Ching Yang^{2,*} and Liang-Yi Li³

¹ Department of Information Communication, Yuan Ze University

² Department of Chinese, National Changhua University of Education

³ Program of Learning Sciences, National Taiwan Normal University

Abstract

This study aims to understand the research trends of artificial intelligence in science education by analyzing publications in three selected journals: *International Journal of Science Education* (IJSE), *Journal of Research in Science Teaching* (JRST) and *Science Education* (SE). The analysis is mainly divided into overall development trends, the trends of empirical research, and issues discussed in non-empirical research. This study analyzed 25 papers after systematic screening. The results showed that the time distribution of these papers roughly reflects the wave of AI. Fifteen of the 25 articles have empirical research. The development trend of empirical research is reported mainly based on the type of artificial intelligence application, the artificial intelligence technology or tools used, research design, research methods, and education level. The results show that profiling and prediction, and assessment and evaluation are two most common applications. Natural language processing and decision trees are the most commonly used artificial intelligence technologies. Experimental design is the most commonly used research design. Traditional testing is the most commonly used research method. Junior high school is the most commonly surveyed education level. Ten of the 25 articles have no empirical research. These 10 articles mainly discuss and report on the following three themes: potential challenges and reflections of artificial intelligence on science education evaluation, potential challenges and reflections of artificial intelligence on science teaching and learning. Potential challenges and reflections of artificial intelligence on science education academic research and journals. Through the analysis and discussion of empirical research and non-empirical research papers, this study finally provides some suggestions for future research and science education.

Key words: Artificial Intelligence, Literature Review, Science Education

* Corresponding author: Hsiao-Ching Yang, jason0209@cc.ncue.edu.tw; ORCID: 009-0002-5239-0322

Received: 2024/4/29, Revised: 2024/7/29, Accepted: 2024/7/30, Available Online: 2024/11/29