

AI 能否成為科學家： 談人工智慧在科學論證中的角色與侷限

吳佳娣¹、劉遠楨²

臺北教育大學¹師資培育處助理教授、²資科系特聘教授兼副校長

什麼是科學推理？

AI 的學習原理又為何？

要想論證 AI 能否成為科學家，首先要先了解科學推理的基本內涵，以及 AI 的學習原理。說明如下：

一、什麼是科學推理？

科學推理 (scientific reasoning) 是一種建立於觀察與實驗基礎上的邏輯思考過程，涵蓋提出假說、進行驗證、分析結果、修正理解等階段，進而逐步形成具備預測與解釋力的理論。這一過程具有動態性與反思性，並強調理論的可驗證性與修正空間。然而，20 世紀重要的科學哲學家 Karl Popper 進一步指出，光是驗證理論的正確性並不足以構成科學，他在《科學的邏輯》(The Logic of Scientific Discovery) 一書中主張：「一個理論若要具備科學性，必須具備可證偽性 (falsifiability)，也就是說，該理論需要能夠被觀察或實驗所推翻 (Popper, 1959)。」Popper 強調，無論累積多少支持證據，都無法證明理論的絕對正確性；相反地，只要有一個能夠被重複觀察的反例出現，即可視為理論的證偽。在此概念下，科學的進步不是靠證明理論得到正確答案，而是藉由發現理論的缺口來促進知識的修正與深化。

在這裡，可能會產生一個常見的疑問：「如果一個理論被證偽了，是不是就表示這個理論從此失效、必須完全拋棄？」Popper 的答案是否定的。被證偽不等於立即作廢，而是代表該理論在特定條件下不再適用，需重新界定其適用範圍或被更完整的理論所取代。舉例來說，一種常見的感冒藥若長期被使用且效果良好，人們便會相信它能有效緩解流鼻水與喉嚨痛等症狀。然而，隨著新型流感病毒的出現，醫師可能發現此藥對某些變異株效果顯著下降，甚至無效。這時候我們不會說這個藥「從此失效」，而會說它在某些條件下被證偽，並需要調整用途或開發替代方案。這正反映出 Popper 所說的「證偽」精神：科學理論的價值並非建基於永不錯誤，而是在於其能夠被挑戰、被修正，並持續接近對自然的理解。因此，真正的科學並非不可動搖的真理，而是一套願意接受反駁與檢驗的知識體系，一個理論若無法說出在什麼條件下會被推翻，它就不具備科學性。

二、什麼是 AI 的學習原理？

那麼 AI 的學習原理是什麼呢？機器學習 (machine learning) 是人工智慧領域中，用以模仿「從經驗中學習」的過程。其核心目標是讓電腦系統在未被明確編程的情況下，透過資料自動學習與推理，從而對新的情境進行預測 (generalization)。以近年被廣泛應用的生成式人工智慧為例，其模型的基本結構如圖 1，表 1 為對應功能的說明。

圖 1
生成對抗網路 (Generative Adversarial Network, GAN) 的結構

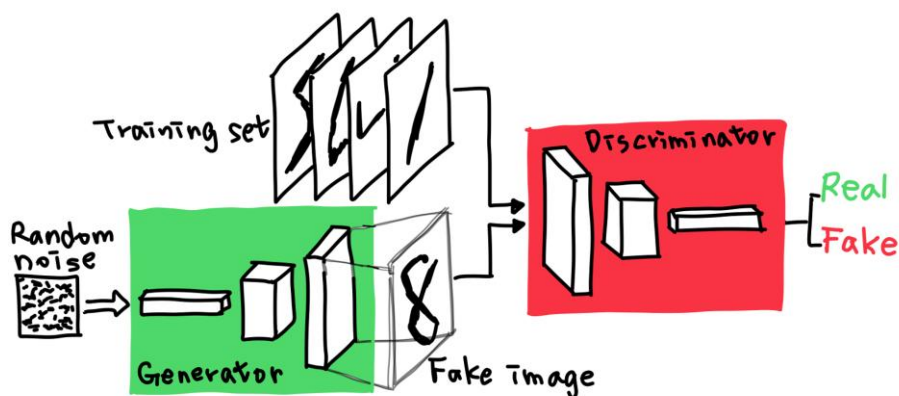


表 1
生成對抗網路 (GAN) 結構對應功能說明

| 區塊 | 名稱 | 功能說明 |
|------------|-------------------------|---|
| 左上 | Training set (訓練資料集) | 提供真實資料大數據給模型學習，讓它知道真圖長什麼樣子。 |
| 左下 (綠色) | Generator (生成器) | 將隨機輸入 (亂數) 轉換為影像的 AI 模型。初始輸出通常缺乏真實性，但隨著訓練不斷優化，可逐步生成高度擬真的影像。 |
| 中間 | Fake image (假圖) | 生成器產出 (假圖) 影像，送由辨別器進行真假判斷。 |
| 右邊 (紅色) | Discriminator (辨別器) | 負責對輸入影像進行二元分類，判斷為真實資料或生成資料，判斷結果同時作為生成器調整參數的重要依據。 |
| 右下 | Real/Fake (判斷結果) | 辨別器的最終輸出，用以標示輸入影像的真實性 (Real 或 Fake)，並將訊號回饋至生成器改進後續生成表現。 |

生成式 AI 的核心功能在於自動生成資料，提供給使用者。其中，最具代表性的架構之一是生成對抗網路（GAN），由 Goodfellow 等人於 2014 年首次提出（Goodfellow et al., 2014）。它由兩個彼此競爭的人工智慧模型組成：生成器（generator）與辨別器（discriminator）。在訓練過程中，首先需要準備大量的真實資料集作為基礎。生成器的任務是將隨機輸入的亂數轉換成「假圖」，而辨別器則負責判斷輸入的影像究竟是真實的資料，還是由生成器產生的偽造圖像。起初，生成器所產生的圖像粗糙且容易被辨別器識破；然而，隨著訓練反覆進行數百萬次，生成器與辨別器皆會不斷調整參數，使其產生的圖像越來越逼真，辨別器也同步提升辨識能力。當這種「生成與對抗」的循環達到平衡時，生成器便能僅憑一組隨機亂數，生成出幾可亂真的圖像，甚至連辨別器也難以分辨其真假，兩個 AI 似乎在玩「騙與抓」的遊戲。

綜上所述，科學推理與 AI 的學習原理雖然都是「從經驗中獲得知識」，但是核心精神存在明顯差異。科學推理的重點在於提出理論並允許理論被挑戰，根據 Popper 的觀點，理論若要具備科學性，就必須能夠在特定條件下被證偽。科學是透過不斷發現反例來修正或取代理論，使知識逐步逼近真實，科學家扮演「主動批判與創造」的角色。人工智慧的學習則是根據大數據與統計方法，以生成對抗網路（GAN）為例，生成器與辨別器透過持續的對抗與迭代訓練，最終能產生高度擬真的結果。

為什麼需要用 AI 介入科學論證？

科學理論之所以具有意義，在於它必須能被檢驗與挑戰。然而要讓理論真正接受檢驗，必須仰賴大量觀測與實驗數據，並且進一步分析。這使得科學家面臨一個新的挑戰：如何在龐大的資訊中快速找到能夠支持或推翻理論的證據？AI 正是在這樣的背景下，於科學推理的過程中扮演關鍵角色：快速處理龐大複雜的資料、揭示人類難以察覺的細節，並且提升研究的效率與準確度。以下舉例說明。

首先，在氣象科學領域，天氣預測向來與人類的生存息息相關。古代社會只能依靠經驗與觀察，例如「早晨有露水，通常白天會是晴天」、「看到螞蟻搬家就會猜測可能要下雨」，準確度有限，且難以應付大規模或極端天氣。即使到了現代，數值天氣預報（numerical weather prediction, NWP）雖然基於嚴謹的物理方程式，但計算成本龐大，且對初始條件極為敏感，一旦資料稍有誤差，預測結果就可能大相逕庭，這使得中長期預報（例如 10-15 天）一直是一項困難挑戰。近期由 DeepMind 團隊開發的 GenCast 系統，它運用生成式機器學習模型進行「機率式天氣預測」，能在短短八分鐘內完成 15 天的全球預報，包含 80 多種氣象變數。研究顯示，GenCast 的表現優於目前世界最先進的歐洲中期天氣預報中心（European Centre for Medium-

Range Weather Forecasts, ECMWF) 的集成系統，特別在極端天氣與颱風路徑的預測上更為精確 (Price et al., 2025)。這意味著，AI 不僅讓天氣預測更快、更準，也為防災決策與能源規劃提供了更可靠的科學依據。

AI 的幫助不只出現在天氣領域，在天文學與醫學中也同樣重要。以天文研究為例，科學家尋找系外行星時，通常依賴觀測恆星光度的細微變化來判斷是否有行星經過。但這些光度資料往往含有大量雜訊，人工檢視既耗時又容易錯過訊號。近年的研究顯示，利用人工智慧中的卷積神經網路 (convolutional neural network, CNN) 分析 NASA Kepler 任務的資料，可以以九成以上的準確率挑選出最有可能的行星候選，顯著提升了研究效率 (Wang, 2025)。同樣地，在醫學領域，大腸鏡檢查雖是大腸癌的重要診斷方式，但小型息肉常因顏色或形態隱蔽而被忽略。He 等人 (2023) 提出的新型 AI 模型 UY-Net，能準確標示息肉，檢測準確率高達 99%，有效降低了誤診的風險。這些案例都說明，AI 的價值不在於取代科學家，而在於協助處理龐大的資料、快速找出關鍵證據，並降低誤差。

其實生活中很多「科普現象」就是 AI 在幫科學家做大數據推理的縮影。例如，手機裡的天氣應用程式，其降雨機率就是 AI 在運算複雜的大氣資料，或者颱風動態的預測；在交通管理上，AI 也能分析市區道路即時流量，協助調整紅綠燈以減少塞車；在環境監測方面，AI 能結合感測器與氣象資料，預測空氣品質與 PM2.5 指數，示警人們是否需要戴口罩或減少戶外活動。這些例子都顯示 AI 不僅改變了科學研究的方式，也融入了我們的日常生活，讓原本抽象的科學論證成果，以具體的應用形式呈現在我們眼前。

如何正確應用 AI 於科學論證？

想像一個場景：某醫院引進了最新的 AI 影像判讀系統，能幫助醫師快速標示 X 光中的異常，但若系統因為訓練資料偏差，把正常組織誤判為病灶，病人可能因此接受不必要的檢查與治療。又或者 AI 天氣預測系統若誤判豪雨，可能導致地方政府做出錯誤的防災決策，例如錯估是否需要停班停課、提前疏散或調度資源，進而造成交通混亂與社會不便。這些例子提醒我們：AI 雖然強大，但若使用不當，反而可能製造新的問題。因此，若要讓 AI 真正成為科學論證的助力，就必須謹慎思考如何正確應用，以下從四個面向來說明：

一、資料品質與偏差控制

AI 的效能高度依賴輸入資料的完整性與多樣性。若資料存在偏差或不足，將直接影響輸出結果，形成「垃圾進、垃圾出」 (garbage in, garbage out) 的

問題。這在天文學研究中尤為明顯，例如利用機器學習分析光變曲線以搜尋系外行星時，若輸入的觀測資料帶有雜訊或樣本不均，模型的結果容易出現誤判或偏差 (Malik et al., 2022)。因此，確保資料品質是科學應用 AI 的基礎前提。

二、可解釋性與透明度

科學的精神在於理論能被檢驗與挑戰，但許多深度學習模型屬於「黑箱」，難以了解其運算邏輯。若一個 AI 系統能準確預測颱風路徑，卻無法說明依據，科學家就無從判斷其可信度。針對此問題，Doshi-Velez 與 Kim (2017) 提出「可解釋人工智慧」(explainable AI, XAI) 的研究框架，強調 AI 模型應具備透明度與可理解性，才能避免與科學方法的核心原則相衝突 (Doshi-Velez & Kim, 2017)。

三、人機協作而非取代

AI 雖擅長於大數據處理與分析，但無法自主提出新問題或建構理論。真正的科學推理仍需人類的批判性思維與創造力。以新藥研發為例，AI 可快速從數百萬種化合物中篩選潛在分子，但最終仍需科學家透過實驗進行生物學驗證，確保結果具科學意義 (Zhavoronkov et al., 2020)。因此，AI 應被定位為輔助者，而非取代人類科學家的角色。

四、道德與責任意識

除了效率與準確性，AI 在科學研究中同樣涉及倫理與責任。若過度依賴 AI，可能導致責任歸屬模糊，例如醫療 AI 誤診導致病患受害時，究竟應由系統開發者、醫師或醫院承擔責任？Floridi 與 Cowls (2019) 提出使用 AI 的五大原則：行善 (beneficence)、不作惡 (non-maleficence)、自主 (autonomy)、正義 (justice)、可解釋性 (explicability)，其中 explicability 特別強調透明與責任歸屬，為未來 AI 與科學結合的重要倫理基礎。

綜上所述，AI 的引入確實大幅提升了科學研究的效率與精準度，並克服了許多傳統方法的限制，使科學論證得以不斷向前推進。然而，若要真正符合科學推理的精神，其應用仍必須遵循四項核心原則：資料品質、可解釋性、人機協作與倫理責任。唯有如此，AI 才能成為推動科學進步的可靠助手，而非隱身於科學背後、操控結論的黑箱之手。正如 Popper 所強調，科學的價值在於能夠被檢驗與挑戰；同樣地，唯有當 AI 的運作與結論也能接受公開檢視與質疑時，它才能真正融入科學論證的核心。

那麼，誰才是真正的科學家？

當我們討論 AI 是否能成為「科學家」時，首先必須釐清科學家的角色。科學家不僅是數據的處理者，更是問題的提出者與理論的建構者。正如 Popper 所言，科學的價值不在於累積支持性的證據，而在於理論能否接受檢驗與挑戰。這意味著，真正的科學精神來自於不斷提出假說、接受反駁，並在錯誤中逐步逼近真理。若從這個角度來看，AI 的強項主要集中在 Battleday 與 Gershman (2024) 所謂的「容易的問題」(easy problems)：這類問題的輸入、輸出與評估標準早已明確界定，因此 AI 能在既定框架內迅速優化，並給出高準確率的答案。

然而，推動科學典範轉移的「困難的問題」(hard problems) 則屬於另一個層次。Battleday 與 Gershman (2024) 指出，真正的科學突破往往不是在既有框架內優化答案，而是徹底重塑問題的定義與概念基礎。典型的例子包括：愛因斯坦 (Albert Einstein) 提出相對論，徹底改寫了時間與空間的概念；孟德列夫 (Dmitri Mendeleev) 建立元素週期表，不僅整理了既有知識，更預測了未知元素的存在；以及拉瓦錫 (Antoine Lavoisier) 透過氧氣理論推翻燃素說，開啟了現代化學革命。這些被視為“hard problems”的經典案例，正凸顯了目前 AI 仍難以企及的領域。

因此，誰才是真正的科學家？現階段的答案仍然是人類。因為唯有人類能將數據轉化為問題、將模式昇華為理論，並在理論遭受挑戰時有勇氣承認錯誤並創造新的框架。總結來說，AI 的確是推動研究不可或缺的助手，能在「容易的問題」上展現驚人的效能；但唯有人類科學家，才能面對「困難的問題」，在批判與想像中引領科學義無反顧地向前，讓科學不只是知識的累積，更是通往更美好生活的希望。又或許真正值得我們思考的，不是 AI 能不能成為科學家，而是當它有朝一日開始像科學家般思考時，我們又該如何重新定義「科學」與「人類」？

參考文獻

- Battleday, R. M., & Gershman, S. J. (2024) . Artificial intelligence for science: The easy and hard problems. *arXiv preprint* arXiv:2408.14508.
<https://doi.org/10.48550/arXiv.2408.14508>
- Doshi-Velez, F., & Kim, B. (2017) . Towards a rigorous science of interpretable machine learning. *arXiv preprint* arXiv:1702.08608.
<https://doi.org/10.48550/arXiv.1702.08608>
- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014) . Generative adversarial nets. *arXiv preprint* arXiv:1406.2661. <https://doi.org/10.48550/arXiv.1406.2661>
- Floridi, L., & Cowls, J. (2019) . A unified framework of five principles for AI in society. *Harvard Data Science Review*, 1 (1) .
<https://doi.org/10.1162/99608f92.8cd550d1>
- He, C. S., Wang, C. J., Wang, J. W., & Liu, Y. C. (2023) . UY-NET: A two-stage network to improve the result of detection in colonoscopy images. *Applied Sciences*, 13 (19) , 10800. <https://doi.org/10.3390/app131910800>
- Malik, A., Moster, B. P., & Obermeier, C. (2022) . Exoplanet detection using machine learning. *Monthly Notices of the Royal Astronomical Society*, 513 (4) , 5505–5516. <https://doi.org/10.1093/mnras/stab3692>
- Popper, K. (1959) . *The logic of scientific discovery*. Routledge.
<https://doi.org/10.1063/1.3060577>
- Price, I., Sanchez-Gonzalez, A., Alet, F., Andersson, T. R., El-Kadi, A., Masters, D., ... Willson, M. (2025) . Probabilistic weather forecasting with machine learning. *Nature*, 637, 84–92. <https://doi.org/10.1038/s41586-024-08252-9>
- Wang, J. (2025) . Training a convolutional neural network for exoplanet classification with transit photometry data. *Scientific Reports*, 15, Article 15408.
<https://doi.org/10.1038/s41598-025-98935-8>
- Zhavoronkov, A., Vanhaelen, Q., & Oprea, T. I. (2020) . Will artificial intelligence for drug discovery impact clinical pharmacology? *Clinical Pharmacology & Therapeutics*, 107 (4) , 780–785. <https://doi.org/10.1002/cpt.1795>