科技文化

科學的科學

——人工智能與邏輯學、語言學、心理學和神經生物學

● 朱嘉明

> 我們的目光所及，只能在不遠的前方，但是可以看到，那裏有大量需要去做的工作。
>
> ——圖靈（Alan Turing）〈計算機與智能〉①

人工智能（AI）逐漸成為現代計算機科學的一個分支，其存在的價值在於揭示智能的實質，並生產出一種新的、能以人類智能相似的方式做出反應的智能機器，使其能夠做出具有主動心智的事情。但是，人工智能思想先於概念，源遠流長，可以追溯到柏拉圖（Plato）、霍布斯（Thomas Hobbes）、萊布尼茨（Gottfried W. Leibniz）和巴貝奇（Charles Babbage），他們已經探討和證明智能不僅體現在人類大腦，也可能發生在人造的機器構造系統之中。

在過去半個多世紀，人工智能完成了從機器學習（machine learning）到深度學習（deep learning）的飛躍，實現了感知、記憶、學習、判斷、推理、行為、情感、心理揣測、意識等認知功能的突破，展現了人工智能存在自我發育和發展，具有認識主體、甚至超越人類的潛質。同時，人工智能研究領域不斷擴大，突破計算機科學的框架，吸納邏輯學、語言學、心理學和神經生物學等學科的學術成果，不僅成為前所未有的跨界科學，而且正在成為「科學的科學」，成為支撐其他科學發展的科學。

一　人工智能與邏輯學

在人類思想史上，邏輯學具有貫穿始終和不可替代的作用。一般來說，邏輯學包括形式邏輯和數理邏輯。形式邏輯，即普通邏輯，研究的是思維規律，包括思維形式及其結構，以及定義、劃分、分析、綜合、試驗、假說等

邏輯方法。亞里士多德(Aristotle)是傳統形式邏輯的奠基人，提出了邏輯思維的三大規律(同一律、矛盾律、排中律)，確定了判斷的定義和分類，制定了演繹三段論(大前提、小前提、結論)推理的主要格式和規則，闡述了演繹與歸納的關係。至於數理邏輯，屬現代形式邏輯，又稱符號邏輯，是基於傳統形式邏輯所演化出來的一門新興的邏輯學科。其特徵是通過數學方法研究關於推理證明等問題，內容包括命題演算、謂詞演算、算法理論、遞歸論、證明論、模型論和集合論等。布爾(George Boole)是數理邏輯的奠基者。

長期以來，人們普遍認為：任何科學都必須遵循邏輯學的基本原理，因為邏輯學對於任何科學都具有普遍適用性。也就是說，科學需要遵循邏輯學的原理，否則會導致科學論證不嚴密。在計算機和人工智能演進過程中，傳統邏輯學是否繼續具有普遍意義呢？這裏需要肯定圖靈和馮．諾伊曼(John von Neumann)的貢獻。誠如論者指出，本世紀中葉，「圖靈作出了偉大的貢獻——將現代邏輯引入了計算機」②。1958年，馮．諾伊曼在他的《計算機和人腦》(*The Computer and the Brain*)中寫道：「我們應該認識：語言在很大程度上只是歷史的事件。人類的多種基本語言，是以各種不同的形式，傳統地傳遞給我們的。這些語言的多樣性，證明在這些語言裏，並沒有甚麼絕對的和必要的東西。正像希臘語或梵語只是歷史的事實而不是絕對的邏輯的必要一樣，我們也只能合理地假定，邏輯和數學也同樣是歷史的、偶然的表達形式」；「人類眼睛上的視網膜，對於眼睛所感受到的視覺，進行了相當的重新組織。這種重新組織，是在視網膜面上實現的；或者更準確地說，是在視覺神經入口的點上，由三個順序相連的突觸實現的；這就是說，只有三個連續的邏輯步驟」，「由此可知，這裏存在着另一種邏輯結構，它和我們在邏輯學、數學中通常使用的理解結構是不同的。前面也講過，這種不同的邏輯結構，其標準是更小的邏輯深度和算數深度(這比我們在其他同樣條件下所用的邏輯深度和算數深度小得多)。因此，中央神經系統中的邏輯學和數學，當我們把它當作為語言來看時，它一定在結構上和我們日常經驗中的語言有着本質上的不同」③。

馮．諾伊曼的思想可以概括為兩點：其一，語言、邏輯和數學都是歷史的，而非永恆的；其二，人工智能的邏輯結構，例如視網膜所體現的邏輯結構，很可能是與邏輯學、數學中通常使用的理解結構不同的「另一種邏輯結構」。在這裏，馮．諾伊曼的思考更多是基於猜想，而人工智能的發展史已經和繼續證明了這種猜想。

人工智能的邏輯和傳統邏輯有所不同，它們存在兩個顯而易見的差別：首先，人工智能邏輯以神經網絡為前提。美國心理學家麥卡洛克(Warren S. McCulloch)和數學家皮茨(Walter H. Pitts, Jr.)並不是簡單地持有一般的唯物主義立場，認為智能是由大腦實現的，他們在〈神經活動中內在思想觀念的邏輯演算〉(“A Logical Calculus of the Ideas Immanent in Nervous Activity”)文章中深刻地指出：「一定類型的(可嚴格定義的)神經網絡，原則上能夠計算一定類型的邏輯函數。」④此說不僅推進了人工神經網絡(Artificial Neural Network, ANN)

的歷史進程，而且為後來1956年的人工智能會議的聯結主義（Connectionism）範式提供了堅實理論基礎⑤。

其次，人工智能邏輯基於機器邏輯推斷模型，即概率圖模型。人工智能的任何邏輯推理的任務，都可以轉換為概率圖模型讓機器進行學習。概率圖模型的數學基礎則是貝葉斯原理（Bayes theorem）和馬爾可夫模型（Markov model）：概率圖分為有向圖（directed graph）和無向圖（undirected graph），包括貝葉斯網絡代表的有向圖模型和馬爾可夫網絡代表的無向圖模型。人工智能通過概率圖模型，即通過有向圖模型和無向圖模型，計算變量間的相關關係或依賴關係，最終實現邏輯推理與歸納。

進一步說，人工智能邏輯體系和自然智能邏輯體系存在如下根本性差異：（1）人工智能試圖找到主體（人或計算機）中的本原元素和邏輯的關係，該主體映射出構成世界的本原客體和它們之間的關係⑥。（2）人工智能邏輯指令系統是「代碼」，「這可以指任何誘發一個數字邏輯系統（如神經系統）並使它能夠重複地、有目的作用的東西」⑦；而自然智能的邏輯系統是自然語言，並不需要代碼。（3）人工智能的數學模型源於大數據統計和概率，自然智能邏輯的數據規模不可比擬。（4）人工智能邏輯是通過機器學習和無限網絡實現的，自然智能邏輯局限於人的大腦和有限網絡來實現。（5）人工智能邏輯的知識系統超越自然智能邏輯的知識系統。（6）人工智能進一步超越了傳統數理邏輯的邊界，既推進了傳統數理邏輯的生產力轉化，也另闢蹊徑用神經網絡實現邏輯再造。（7）人工智能邏輯超越傳統邏輯歸納和演繹的功能，模糊歸納和演繹的界限。這是因為人工智能技術最基本的兩種能力就是歸納和演繹能力。機器學習就是一種典型的歸納方法，深度學習則根據已有的知識來進行邏輯推理和計算，實現演繹。

特別需要強調，並非思維邏輯必須受制於歸納和演繹模式。愛因斯坦（Albert Einstein）對想像和思辨有很高的評價，提出概念是思維的自由創造，科學史不乏「不能歸納得到」的事例⑧。理論和經驗不可分割，所以人工智能的邏輯是超越傳統邏輯框架和結構的。

近年來，各類大語言模型（Large Language Model, LLM）的產生，本質上就是人工智能邏輯模型，實現了數字方法與形式邏輯和數理邏輯、歸納和演繹功能的集大成。自此，邏輯學分類發生改變：進入自然智能邏輯和人工智能邏輯並存的歷史階段。

二　人工智能與語言學

語言學是以人類語言為研究對象，探索範圍包括語言的性質、功能、結構、運用和歷史發展，以及其他與語言有關問題的學科。人類語言具有以下特徵：（1）世界的存在體現為語言的存在，語言和世界是同構的。所以，維特根斯坦（Ludwig Wittgenstein）認為：語言的邊界就是世界的邊界⑨。（2）語言

系統與符號系統等價。(3) 語言是人類思維的基礎，語言是自然智能的載體。(4) 人類知識即是語言的組合和序列。(5) 語言具有運算和程式的天然功能。(6) 語言在很大程度上只是歷史的事件⑩。

在現代語言學領域，維特根斯坦的地位是不可逾越的，他的劃時代貢獻是為語言劃定了範圍：語言屬經驗世界，即可以言說和可以說清楚的問題；在經驗世界的範圍之外，則是不能夠言說和說清楚的問題，例如形而上學、倫理學和美學。他特別強調的是，並不存在唯一的「反映世界」的語言規則，語言如同遊戲，不同的遊戲具有不同的規則。他將其觀點總結為「語言遊戲」(language game) ⑪。

在1930年代末，近五十歲的維特根斯坦和年輕幾歲的圖靈都在劍橋大學教書，分別開設了一門名叫「數學基礎」的課程。維特根斯坦教授的「數學基礎」的核心思想是：數學是發明而不是發現，數學命題僅僅是一種語法，是人類的一種語言結構；維特根斯坦關注的是語言。圖靈的「數學基礎」的核心思想是：數學通過選取一組嚴密而簡潔的公理，實行一種邏輯證明；圖靈關注的是計算。

維特根斯坦有句名言：如果獅子會說話，「我們就不會理解牠」⑫。因為獅子所嵌入的「生命形式」，完全不同於人的「生命形式」。語言是為它所產生的生命形式服務，反過來也塑造了語言。進一步說，即使獅子能說話，牠理解世界的方式也會與人作為物種的思維方式不同，以至於人最終不會理解獅子。按照這個邏輯，計算機與人的差別也是如此，因為機器與人類存在不同的「生命形式」。所以，維特根斯坦認為圖靈1936年的文章〈論可計算數，及其在判定問題中的應用〉(“On Computable Numbers, with an Application to the Entscheidungsproblem”)，「那不過是人在計算而已」⑬。

維特根斯坦顯然低估了圖靈1936年文章的意義：從本質上說，圖靈機是一種抽象的計算機模型(下詳)，實現在任何可計算的範疇內的計算問題。如果智能是可以界定和量化的行為，智能是可以計算的，計算就是一種智能形態。著名的「邱奇—圖靈論題」(Church–Turing thesis) 可以表述為：所有計算裝置和人按照算法執行的計算和圖靈機等價，最終人的智能和圖靈機的能力等價。

人工智能演變的歷史似乎證明圖靈的觀點很可能是正確的。計算機、人工智能就類似獅子，不僅創造了不同的「生命形式」和語言存在，而且可以與人的智能等價。可計算和量化的智能實現了對自然語言的超越，不僅如此，也不再存在「自我」與「非自我」等知覺概念的本質差別。機器可能「思考」，至於機器是否可能理解自己，並非那麼重要。如同人可以思考，並不等於人可以完全理解人本身一樣。當然，必須承認：儘管人工智能和自然智能具有可以溝通的功能，卻有着不同特徵的語言體系。1949年，數學家香農 (Claude E. Shannon) 與韋弗 (Warren Weaver) 提出香農—韋弗模式 (Shannon–Weaver Model)，即通過機器實現自然語言處理 (Natural Language Processing, NLP) 的原理⑭。

1980年，美國著名哲學家塞爾(John R. Searle)在〈心靈、大腦與程式〉("Minds, Brains, and Programs")一文中⑮，提出名為「中文房間」的思想實驗：假想一位只會說英語的人，身處於一個只有一扇小窗口的封閉房間，房間有足夠的稿紙和筆。他隨身帶着一本寫有中文翻譯程式的書。房間外的人將寫着中文的紙片通過小窗口送入房間，房間內的人通過那本寫有中文翻譯程式的書，翻譯這些文字，並用中文回覆。雖然房間裏的人不懂中文，外面的人卻以為裏面的人懂中文。這個「中文房間」思想實驗，是為了推翻圖靈測試(The Turing Test)，證明計算機足以運行一個程式，處理信息，然後給出一個智能的印象。「中文房間」的思想實驗涉及到語義語法、意向性、身心等問題，要反駁這個思想實驗並非困難。不過，我們不能因為塞爾不理解中文，而得出「中文房間」這個運行程式沒有理解能力的結論。論者指出，「即使最簡單的程式也並不是純形式主義的，而是具有某種相當本原的語義特性，所以從根本上說，計算理論並非不能解釋意義」⑯。

自然智能的載體是人本身，即是一台生物機器，也是一台語言機器，是人類長期進化的結果。人類基因的本質就是程式化的代碼語言。至於與計算機結合的人工智能，同樣也是一台語言機器，只是最初的計算機比特(bit)語言是人類創造的、軟件需要與硬件支持的機器結合。所以，人工智能就是機器智能。

如何處理人類自然語言從來是人類智能的標誌性能力。到現在為止，計算機科學和人工智能科學的歷史證明了三點：其一，計算機的比特語言和自然語言相比，在實現信息儲存、交易和推理方面更有效率。其二，人工智能語言可以解決維特根斯坦所說的不能夠言說和不能夠說清楚的問題。因為人工智能語言的程式，打破了自然語言的局限性。其三，未來的語言學將包括自然語言和人工智能語言。人工智能語言可能兼容自然語言，反之則不可能。也就是說，人工智能的生命形態可以擺脫生物化學的限制，具有更快的進化速度。

2022年美國人工智能研究公司OpenAI推出的聊天機器人程式ChatGPT所代表的各類大語言模型的根本意義在於：輸入的是自然語言，中間過程是機器語言，輸出的是自然語言，實現了自然語言與人工智能語言的融合。只是在這樣的歷史過程中，人工智能語言和自然語言將進入失衡狀態：人工智能可以讀懂自然語言，自然智能則不需要理解機器語言。語言史表明，語音文字和抽象文字發源於以象形文字代表的形象語言形態。

大語言模型具備海量的語言數據庫，通過前所未有的算力，對語言進行無限的計算、分析和組合，重新構建和實驗全新的語言體系。大語言模型拉開了人工智能主導的語言學革命序幕，人工智能的進展和語言學革命的進展，進入互動的全新的歷史階段。從哲學高度看：「大語言模型的勝利並不完全是一種經驗主義的勝利，語言的結構並非單純是經驗的。用哲學的行話說，大語言模型是一種理性主義和經驗主義結合的勝利。」⑰

人類文明的早期記錄和兒童對世界的認知，都是從依賴直觀形象和表象的形象思維開始的。2024年2月，OpenAI發布的文字生成視頻模型Sora，顯

示了人工智能可以實現從文本語言轉化為圖像和視頻語言，進而更真實地模擬部分的物理世界。以Sora為先鋒的視覺大模型，本質就是多模態語言模型。通過Sora，發生人類思維模式基於人工智能技術的一種回歸：通過人工視頻的形象思維—抽象語言思維—物理世界的形象思維。

三　人工智能與心理學

人的自然智能和心理學是不可分割的。那麼，人工智能科學與心理學的關係如何？在圖靈看來，不僅機器可以思維，而且可以建構反映人類詳盡心理過程的計算機模型。針對圖靈的主張，反對意見認為：「要使計算機的表現在深度、廣度以及靈活性上與人類心智相媲美，在原理上和(或)實踐上，都是不可能的。一台閱讀十四行詩的計算機，不管是真有智能，或者僅僅只能模仿智能，都決無存在的可能。」⑱以上分歧涉及了人工智能科學和心理學的關係。按照圖靈的想法，人類的心理活動是可以最終模型化和計算的；反對的意見則認為人類的心理活動是沒有可能被計算的。

心理學是研究心理現象發生、發展和活動規律的科學。心理學可以分為基礎心理學和應用心理學。基礎心理學包括：認知；情緒、情感和意志；需要和動機；能力和人格。其中的認知涵蓋感覺、知覺、記憶、思維等心理現象。不論人類心理活動如何複雜，從根本上說，都是人的神經系統運行的結果。神經系統的複雜程度與智能程度是正相關的。換一個角度說，神經系統是心理現象產生的物質基礎，心理活動是神經系統機能的反映。所以，只要人工智能技術和人的神經網絡結合，就意味着可以實現人的認知活動、進而實現人的心理活動的可計算。

如果進一步研究十九世紀以來心理學的主要流派，例如「構造心理學」、「機能心理學」、「行為主義心理學」、「整體心理學」、「認知心理學」、「生理心理學」，還有「精神分析」，不難發現其共同特徵：人類的一切心理活動和現象都是可以分解和解構的，都是可以量化的。這些心理學流派代表，或者沒有看到，或者沒有認識到，計算機和人工智能的思想和技術有助於驗證他們的理論。在人工智能科學和心理學結合的過程中，最終推動了心靈計算理論(Computational Theory of Mind, CTM，又稱「心智計算理論」)的形成。

在這個領域，1936年圖靈機的誕生具有劃時代的意義。圖靈並非心理學家，他認為機器具有學習能力，進而可以思想，再進入到心理層面。圖靈機是一種無限記憶自動機，圖靈的基本思想就是用機器來模擬人們用紙筆進行數學運算的過程，「每個算法可在一台圖靈機上程式化」⑲。1943年，麥卡洛克和皮茨在前述的〈神經活動中內在思想觀念的邏輯演算〉一文中，最早提出神經活動具有計算性，以及認知可以由計算來解釋，並基於數學和一種稱為「閾值邏輯」的算法創造了一個非常簡單的神經元模型，即M-P模型(又稱「麥卡洛克—皮茨模型」或「MCP模型」)，成為神經網絡模型的理論研究的拓荒者

和先驅。該模型將神經元當作一個功能邏輯器件來對待⑳。之後，這種模型使得神經網絡的研究分裂為兩種不同研究思路：一種主要關注大腦中的生物學過程，另一種主要關注神經網絡在人工智能裏的應用。自此，神經網絡研究奠定了心靈計算理論的前提：心智／大腦是一個信息處理系統、一個計算系統，而思維是一種計算形式。更深層的思想是：人類的一切精神現象，包括情感，都可以被認定為「信息」。既然是信息，就可以模型化。

心靈計算理論的形成，直接推動了計算主義（Computationalism）和認知計算主義（Cognitive Computationalism）的出現。計算主義的基本思想是，心理狀態、心理活動和心理過程是計算狀態、計算活動、計算過程。簡言之，認知就是計算。徹底的計算主義者認為：計算主義就是世界觀。二十世紀80年代以後，認知科學發生了一場「人工神經網絡革命」，認知科學的聯結主義研究範式取代了符號主義（Symbolism）範式。聯結主義的核心思想就是一切人類認知活動完全可歸結為大腦神經元的活動。

討論計算主義，派利夏恩（Zenon W. Pylyshyn）是一位繞不過去的重要人物，代表作是《計算與認知》（*Computation and Cognition: Toward a Foundation for Cognitive Science*）㉑。該書的核心思想是：人就是一部邏輯機器，認知的本質是計算，一切認知過程和智能行為都是可計算的。派利夏恩認為，「對心理狀態的語義內容加以編碼通常類似於對計算表徵的編碼」。也就是說，人類以及其他智能體實際上就是一種認知生靈、一種計算機，這就是著名的「計算機隱喻」：「這種思想基於認知科學的一個中心假定：對心智的最恰當理解是將它視為心智中的表徵結構以及在這種結構上操作的計算程序。這種心智的計算觀假定心靈具有心理表徵，它類似於計算機器的數據結構，而心靈活〔動〕中的計算程式類似於計算機器的算法。這樣一來，我們的思維或心智就類似於計算機的運行程度了。」㉒

紐厄爾（Allen Newell）和西蒙（Herbert A. Simon）對於計算機和心靈的關係有着更為肯定的觀點：「心靈是一個計算系統，大腦事實上是在執行計算的職能（計算對智能來說是充分的），它與可能出現在計算機中的計算是完全等同的。」而對於紐厄爾和西蒙觀點的批評者來說，人工智能並非不能實現，「而是認為這比起他們兩位定義中提出的那種文字形式要複雜得多」㉓。

技術層面的認知計算（cognitive computing）是一種新型的計算，其目標是建立關於人腦／大腦如何感知、推理和響應刺激的更精確模型。認知計算具象為基於人工智能和信息科學的技術平台。這些平台包括機器學習、推理、自然語言處理、語音識別和視覺的對象識別、人機交互、對話和敘述生成等技術。

值得重視的是，伴隨人工智能愈來愈強悍的自我發展和超級人工智能的成熟，人工智能勢必形成自己的心理系統。於是，未來將出現兩種心理體系：人類的和被人工智能化的心理體系；人工智能自身發展出來的人工智能心理體系。智能心理學（或稱「AI心理學」）將會產生。科幻小說家阿西莫夫（Isaac Asimov）的預言離我們愈來愈近。例如，現在的機器人夥伴，具有共情和機器人情感。共情部分基於感知的情感模型，根據感官信息了解人的情感

狀態；而通過一種循環脈衝神經網絡來改進機器人的情感模型，是利用人的情感信息、機器人的內部信息和外部信息計算出機器人的情感狀態。機器人夥伴可以利用情感結果來控制面部和手勢表情，言語風格也會隨着機器人的情緒狀態而改變，能夠實現機器人夥伴與人類進行情感和自然的互動。不過，現在還難以預見的是，這個世界是否最終可以進入人工智能的情感處於絕對主導的時代。

若要討論計算主義，我們難以迴避哥德爾（Kurt F. Gödel）的「不完備定理」（Incompleteness Theorem）。「不完備定理」並沒有設定人類理性極限，只是揭示了數學形式主義的內在矛盾㉔。由該定理引申：在機器模擬人的智能方面存在着某種極限，或者說計算機永遠沒有可能做到人所能做的一切。也許可以說，哥德爾不排除存在事實上等價於數學直覺的定理證明機器，但是這不意味着人類可以確切知道它是否會準確無誤地工作。

至於圖靈提出的「心智過程不能超越機械過程」的論斷，哥德爾認為需要兩個假設：其一，沒有與物質相分離的心；其二，大腦的功能基本上像一台數字計算機。後者的概率很高，而前者將是被科學否定的時代偏見。此外，還取決於包括大腦生理學在內的整個科學的進展。從新近發現的哥德爾的一部分重要手稿和1970年代與王浩的談話記錄中我們得知，哥德爾在嚴格區分心、腦、計算機的功能後，明確反對「心腦同一論」：「心與腦的功能同一卻是我們時代的偏見。」一方面，哥德爾承認「大腦的功能不過像一台自動計算機」，大腦的計算主義成立；另一方面，哥德爾卻懷疑和否定可以計算心的活動，因為沒有足夠的大腦神經元來實現心的複雜運作。哥德爾曾這樣解釋「心」的含義：「我所說的心是指有無限壽命的個體的心智，這與物種的心智的聚合不同。」人心有洞察具有超窮性質的數學真理的直覺能力，特別是能夠洞察數學形式系統的一致性㉕。

歷史終於到了一個十字路口：人工智能是否可以實現心、腦、計算機的統一？大模型是否就是一個證明？如果是這樣，意味着儘管哥德爾「不完備定理」繼續存在，但是他否定心和心理活動的可計算的邏輯不再成立。未來很可能是這樣：人工智能可以幫助我們認識心靈。正如丹尼特（Daniel C. Dennett）指出，雖然「AI尚未揭開古老的心靈之謎，但是它為我們提供了規範和拓寬哲學想像力的新方法，至於對這些方法的利用，我們還剛剛開始」㉖。

四　人工智能與神經生物學

神經生物學的思想歷史悠久，但是直到十九世紀末，因為顯微鏡等技術的發展，生物學家才開始認知神經元的結構和功能。1888年，西班牙神經解剖學家卡哈爾（Santiago Ramón Cajal）提出了神經元學說，即神經系統是由一個個神經元構成的，奠定了現代神經生物學的理論基礎。1897年，謝靈頓（Charles S. Sherrington）提出使用「突觸」（synapse）這個術語來描述一個神經元

與另一個進行信息溝通。二十世紀初，英國生理學家艾略特(Thomas R. Elliott)和戴爾(Henry H. Dale)提出了突觸傳遞(synaptic transmission)，也就是神經元之間的通訊是通過突觸實現的假說。之後，這一假說得到實驗證實，確定神經信號傳遞的模式。到了二十世紀60年代，「樹突」(dendrites)的整合功能被認知，證明神經系統有無衝動的突觸迴路和突觸相互作用。神經生物科學形成了以下關鍵性共識：神經元的核心特點是多輸入、單輸出；突觸兼有興奮和抑制兩種性能，即能時間加權和空間加權，可產生脈衝；脈衝進行傳遞，且非線性。

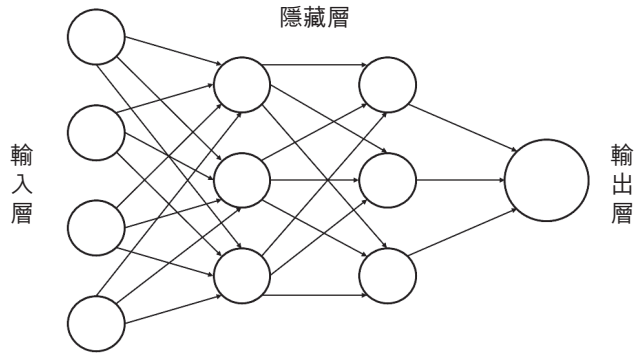
神經生物學的發展，對神經元、突觸、信號傳遞、腦活動等方面的研究與試驗的突破，不僅極大地推動了生命科學和醫學的發展，而且對人工智能科學產生靈感、啟發和持續影響。在現代計算機發展的早期階段，人工智能、神經科學和心理學領域的科學家已經開始了跨學科研究。

馮．諾伊曼在《計算機和人腦》的第二部分「人腦」中，在肯定大腦仍然是唯一已知的真正通用智能系統前提下，詳盡地分析了現代計算機和人類神經系統這兩類「自動機」之間的相似和不同之處，做了「神經元的大小、它和人造元件的比較」，探討了「神經系統內的記憶問題」，描述了「神經系統中記憶容量的原理」。他特別強調的是：「對神經系統作最直接的觀察，會覺得它的功能顯而易見地是數字型的」，「神經系統中使用的信息系統，其本質是統計性質的」。神經系統中的動態過程特徵，從數字的變為模擬的，從模擬的又變回來成為數字的，不斷反覆。也就是說，神經系統的信息並非是規定的符號，也不是精確的數字，而是表現為周期性或近似周期性的脈衝序列的頻率，「所以，完全有理由設想，這些脈衝序列之間的一定的(統計的)關係，是可以傳送消息的」[27]。

馮．諾伊曼的思想邏輯堪稱完善：其一，神經系統屬數字型和具有統計性質的系統，神經系統包含數字部分和模擬部分，所以，神經系統可以轉化為基於統計方法的神經模型。其二，神經系統可以通過物理化元件和軟件形態的代碼，實現算數深度和邏輯深度的相互轉化，構建現代模擬計算機，即賦予人工智能的計算機。人工智能歷史正是沿着這個思路發展的。至少一部分人工智能科學家相信：通過構建類似於人腦中神經元之間相互連接的網絡結構，即構建人工神經網絡。人工神經網絡可以定義為：一種通過模擬人的大腦神經結構去實現人腦智能活動功能的信息處理系統，基於模仿大腦神經網絡結構和功能的數學模型，是人腦的一種抽象、簡化和模擬模型，能夠進行複雜的邏輯操作和實現非線性關係的信息處理系統。人工神經網絡是一種非線性、統計性數據建模工具，通常採用拓撲結構。

人工神經網絡的物理特徵就是一個由大量簡單元件相互連接而成的複雜網絡。它由一定數量的基本單元分層聯接構成，每個單元的輸入、輸出信號以及綜合處理內容比較簡單，網絡的學習和知識存儲體現在各單元之間的聯接強度上，每個節點執行一個簡單的計算。下圖是人工神經網絡體系的示例(圖1)。

圖 1　人工神經網絡體系



圖片來源：筆者改製自Moonzarin Reza, "Galaxy Morphology Classification Using Automated Machine Learning", *Astronomy and Computing*, vol. 37 (October 2021), https://doi.org/10.1016/j.ascom.2021.100492。

在過去的半個多世紀，巨量的智慧、人才和資本投入到人工神經網絡的研究和開發之中，其發展呈現高潮、低潮和再高潮。人工神經網絡歷史的主要事件包括：1980年，出現卷積神經網絡（Convolutional Neural Network, CNN）的雛形「新認知機」（Neocognitron）。1984年，辛頓（Geoffrey E. Hinton）與謝諾夫斯基（Terrence J. Sejnowski）等合作提出了大規模並行網絡學習機和隱藏單元的概念，這種學習機後來被稱為「玻爾茲曼機」（Boltzmann Machine）。1986年，魯姆哈特（David E. Rumelhart）、辛頓等人發明了適用於多層感知器（Multilayer Perceptron, MLP）和誤差反向傳播（backpropagation）的算法，並採用S型函數（sigmoid function）進行非線性映射，解決了非線性分類和學習的問題。1989年，海希特－尼爾森（Robert Hecht-Nielsen）證明了多層感知器的萬能逼近定理；同年，楊立昆（Yann LeCun）發明卷積神經網絡LeNet-5，並用於數字識別。這一年，Q學習（Q-Learning）算法問世，它是一種無模型強化學習（reinforcement learning）算法。1991年，循環神經網絡（Recurrent Neural Network, RNN）出現。1992年，Q學習的收斂性被證明。1997年，長短期記憶網絡（Long Short-Term Memory, LSTM）被提出。1987年，首屆國際神經網絡大會在聖地亞哥召開，國際神經網絡協會（International Neural Network Society）成立。客觀地說，到1990年代，多層神經網絡研究已經日趨成熟，只是那個時期的算力太小，大數據規模有限，難以得到廣泛應用。

人工神經網絡發展歷史的重大拐點是2006年：辛頓和他的學生薩拉赫丁諾夫（Ruslan R. Salakhutdinov）在《科學》（*Science*）雜誌發表題為〈使用神經網絡降低數據維度〉（"Reducing the Dimensionality of Data with Neural Networks"）的論文，正式提出了「深度學習」概念[28]。在這一年，「深度置信網絡」（Deep Belief Nets/Deep Belief Network, DBN）、堆疊自編碼器（Stacked Autoencoder）和連結時序分類（Connectionist Temporal Classification, CTC）相繼問世，深度卷積網絡LeNet-5獲得訓練突破。特別值得注意的是，人工神經網絡引入圖形

處理器（Graphics Processing Unit, GPU）。GPU是一種特殊類型的處理器，在具備大量重複數據集運算和頻繁內存訪問等特點的應用場景中，特別是圖形處理上，GPU相比中央處理器（Central Processing Unit, CPU）具有顯著優勢。它為深度學習提供了堅實的硬件支持。

深度學習雖然是機器學習的一種，但是兩者之間存在實質性的差別。深度學習本質上就是一種人工神經網絡。人工智能完成從機器學習到深度學習的飛躍，使之和其他人工神經網絡緊密互動，甚至一體化。最終人工智能發展主要依靠的是深度學習。

「深度學習」的概念和「人工神經網絡」的概念不可分割。不存在沒有人工神經網絡為基礎的深度學習，也不存在沒有深度學習能力的人工神經網絡。深度學習結構包含多個隱藏層的多層感知器。深度學習通過組合「低層」特徵形成更加抽象的「高層」屬性類別或特徵，以發現數據的分布式特徵表示。深度學習的動機在於建立模擬人腦進行分析學習的神經網絡，模仿人腦的機制來解釋數據，例如圖像、聲音和文本等[29]。深度學習理論和模式，開啟了深度學習在學術界和工業界的浪潮。2011年之後，微軟（Microsoft）首次完成語音識別的重大突破。

人工神經網絡已經發展成為一個體系。深度神經網絡模型主要有卷積神經網絡、遞歸神經網絡（Recursive Neural Network）、深度置信網絡、深度自編碼器（Autoencoder）和生成對抗網絡（Generative Adversarial Network, GAN）等。以深度學習和卷積神經網絡的關係為例，卷積神經網絡的研究始於二十世紀80至90年代，包含卷積計算且具有深度結構的前饋神經網絡（Feedforward Neural Network, FNN），類似於人工神經網絡的多層感知器。進入二十一世紀後，因為深度學習理論的提出和數值計算設備的改進，卷積神經網絡作為一種深度學習模型和深度學習的一種代表算法，終於得到普遍認知，2012年之後卷積神經網絡在圖像分類、圖像分割、目標檢測、圖像檢索等領域處於主導地位，並蔓延到計算機視覺、自然語言處理等領域。

因為人工神經網絡和深度學習的融合，催生了各類大語言模型的出世。2015年，美國斯坦福和柏克萊大學四名學者聯名發表了基於深度學習方法、結合了卷積神經網絡和生成對抗網絡的技術，主要用於圖像生成的擴散（Diffusion）模型[30]。2017年，谷歌（Google）團隊發布的論文〈注意力足矣〉（"Attention Is All You Need"）是里程碑事件[31]。該論文提出了一種新的網絡架構，稱為"Transformer"[32]，它僅基於「注意力機制」（attention mechanism），可以不需要循環或卷積支持。Transformer已被證明在質量上更優越，在並行性和訓練速度方面也比傳統序列轉換模型更快。2024年2月，OpenAI發布的Sora即基於和疊加了擴散模型和Transformer模型的各自優勢。

2019年，谷歌發布BERT語言模型。BERT代表來自Transformer的雙向編碼器表示（Bidirectional Encoder Representations from Transformers）。之後，眾多大語言模型相繼出現。2022年，ChatGPT所代表的大語言模型獲得前所未有

的影響力。大語言模型，說到底就是神經網絡模型的子集。大語言模型的主要特徵是呈複雜結構，通常由多個隱藏層組成，每個隱藏層包含大量的神經元，並且具有數百萬到數十億甚至更多的可訓練參數[33]。

當然，不論是神經網絡大模型還是大語言模型的運行，最終還是依賴物理資源。語言訓練大模型尤其需要使用高性能計算設備（如GPU、TPU〔Tensor Processing Unit，張量處理器〕）或雲計算平台的支持。所以，1993年在美國加州成立的以銷售GPU為主的無廠半導體公司輝達（NVIDIA）得以崛起。

五　結語

至此，本文討論了人工智能與邏輯學、語言學、心理學和神經生物學的關係，以此證明人工智能具有「科學之科學」的特徵。在人工智能，包括大模型的成長過程中，發生了人工智能和信息科學形成過程中從邏輯到經驗、從理論到技術的相互影響和融合的現象。信息技術和人工智能技術與計算機科學的深厚淵源，導致兩者之間不僅沒有清晰的界限，相反，卻具有顯而易見的相似性。

二十世紀80年代，人們關注到在信息領域的數據—信息—知識三元組模型（圖2a）；後來，這一模型擴展至四個層次，頂層為智慧（圖2b）。

圖2a　信息領域的三元組

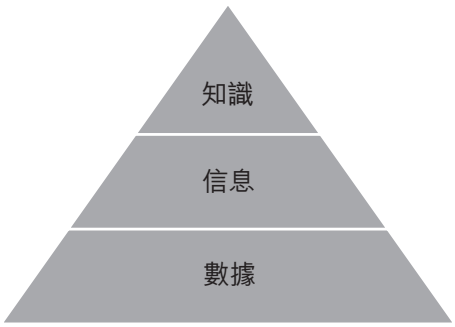
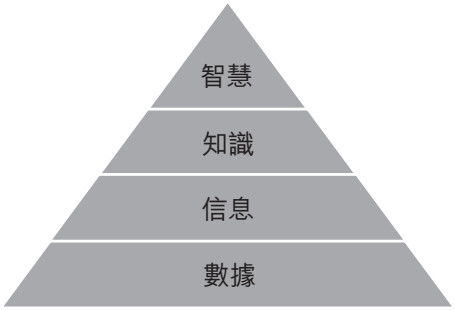


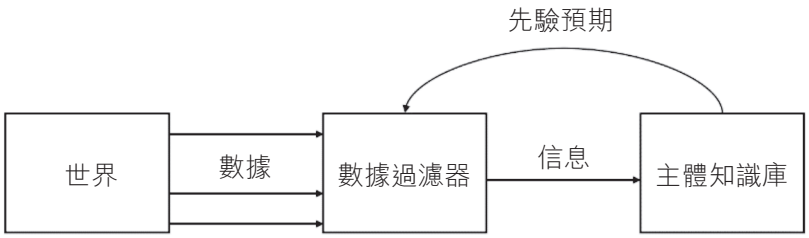
圖2b　信息領域的四個層次



圖片來源：筆者繪製。

圖2a和2b的最底層是數據，數據是資源，具有可操作性，還可以被度量。度量數據最流行的方法是按二進制位比特。信息是具有意義和重要性的符號集合。至於信息和知識的關係是複雜的，信息可以產生知識，信息也可以從知識產生數據。知識被視為一種信息完成行為，獲取知識意味着不確定性的減少。布爾金（Mark Burgin）在《信息論：本質．多樣性．統一》（*Theory of Information*: *Fundamentality, Diversity and Unification*）中，引用了布瓦索（Max H. Boisot）的一個「數據到知識的轉變」示意圖（圖3）。

圖3　數據到知識的轉變



圖片來源：布爾金（Mark Burgin）著，王恆君、嵇立安、王宏勇譯：《信息論：本質・多樣性・統一》（北京：知識產權出版社，2015），頁116。

無論如何，知識既是自然智慧，也是人工智慧的基礎，所以圖2b與人工智能構造近似。因為從本質上說，人工智能大模型就是一個不斷輸入數據、數據轉化為信息、輸出不同形態（多模態）新知識的系統。

註釋

① A. M. Turing, "Computing Machinery and Intelligence", *Mind* LIX, no. 236 (1950): 460.

② 紐厄爾（Allen Newell）、西蒙（Herbert A. Simon）：〈作為經驗探索的計算機科學：符號和搜索〉，載博登（Margaret A. Boden）編，劉西瑞、王漢琦譯：《人工智能哲學》（上海：上海譯文出版社，2001），頁176。

③⑦⑩㉗ 馮・諾意曼（John von Neumann）著，甘子玉譯：《計算機和人腦》（北京：商務印書館，1979），頁59-60；51；59；29-58。

④⑯⑱㉓㉖ 〈導言〉，載《人工智能哲學》，頁4；8；8-9；11-12；27。

⑤ 關於人工智能的符號主義、聯結主義與行為主義，參見Melinda Bognár, "Prospects of AI in Architecture: Symbolicism, Connectionism, Actionism" (10 January 2021), https://openreview.net/pdf?id=gvHffM4DlpG。

⑥ 休伯特・德福雷斯（Hubert L. Dreyfus）、斯圖爾特・德福雷斯（Stuart E. Dreyfus）：〈造就心靈還是建立大腦模型：人工智能的分歧點〉，載《人工智能哲學》，頁421。

⑧ Albert Einstein and Leopold Infeld, *The Evolution of Physics* (New York: Simon and Schuster, 1966).

⑨ Ludwig Wittgenstein, *Tractatus Logico-Philosophicus*, trans. C. K. Ogden (London: Kegan Paul, Trench, Trubner & Co., Ltd., 1922), 74.

⑪ Ludwig Wittgenstein, *Philosophical Investigations*, trans. G. E. M. Anscombe (Oxford: Basil Blackwell, 1958).

⑫ Ludwig Wittgenstein, *Philosophical Investigations*, 223.

⑬ 論者指出：「現在的證據表明維特根斯坦在1939年時已讀過圖靈1936的文章，並且有評論：『那不過是人在計算而已。』」參見〈對掐：維特根斯坦和圖靈〉（2022年8月16日），搜狐網，www.sohu.com/a/577150086_121124776。圖靈文章參見A. M. Turing, "On Computable Numbers, with an Application to the Entscheidungsproblem", *Proceedings of the London Mathematical Society* s2-42, issue 1 (1937): 230-65。

⑭ Claude E. Shannon and Warren Weaver, *The Mathematical Theory of Communication* (Urbana, IL: University of Illinois Press, 1964).
⑮ John R. Searle, "Minds, Brains, and Programs", *Behavioral and Brain Sciences* 3, no. 3 (1980): 417-24.
⑰ 梅劍華：〈人工智能會重塑哲學嗎？〉，《信睿周報》，第113期（2024年2月1日），https://mp.weixin.qq.com/s/ZwWi3FRfK0BXGsdtuBy48w。
⑲ David Harel, *Algorithmics: The Spirit of Computing*, 2d ed. (Reading, MA: Addison-Wesley, 1992), 233.
⑳ Warren S. McCulloch and Walter Pitts, "A Logical Calculus of the Ideas Immanent in Nervous Activity", *Bulletin of Mathematical Biology*, vol. 5 (December 1943): 115-33.
㉑ Zenon W. Pylyshyn, *Computation and Cognition: Toward a Foundation for Cognitive Science* (Cambridge, MA: MIT Press, 1984).
㉒ 任曉明、胡寶山：〈為認知科學的計算主義綱領辯護——評澤農．派利夏恩的計算主義思想〉，《江西社會科學》，2007年第2期，頁46、48。
㉔ Kurt Gödel, "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I", *Monatshefte für Mathematik und Physik* 38, no. 1 (1931): 173-98.
㉕ 柳樂：〈哥德爾與人工智能〉（2012年6月30日），www.douban.com/note/222807957/?from=gallery_new_post&_i=8142413TSlQ7hK。
㉘ G. E. Hinton and R. R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks", *Science* 313, no. 5786 (2006): 504-507, https://doi.org/10.1126/science.1127647.
㉙ 深度學習通過無監督的學習方法逐層訓練算法，再使用有監督的反向傳播算法進行調優，用以解決「梯度消失」（gradient vanishing）問題。所謂「梯度消失」是指在深度學習模型訓練過程中，參數的梯度值變得非常小，接近於零，導致模型參數更新緩慢，從而影響模型的訓練效果。
㉚ Jascha Sohl-Dickstein et al., "Deep Unsupervised Learning Using Non-equilibrium Thermodynamics", *Proceedings of Machine Learning Research*, vol. 37 (2015), https://proceedings.mlr.press/v37/sohl-dickstein15.pdf.
㉛ Ashish Vaswani et al., "Attention Is All You Need" (2017), https://papers.nips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.
㉜ 中文將"Transformer"翻譯為變換器，並不能完全反映大模型的Transformer的基本內涵。所以，本文還是直接使用英文原詞。
㉝ 自OpenAI在2022年11月發布ChatGPT，僅兩個月就有一億用戶參與，成為有史以來用戶增長最快的產品。之後，全球範圍內開始了「百模大戰」。OpenAI推出了GPT-4、GPT-4V和ChatGPT-4，並圍繞着ChatGPT推出了ChatGPT plugins、Code Interpreter、GPT Store、ChatGPT Team等。微軟推出了Bing Chat（後來改名為Copilot）、Copilot for Microsoft 365等產品。谷歌推出Bard（後來改名為Gemini）。Meta推出了LLaMA、LLaMA2等。馬斯克（Elon Musk）成立的X.ai公司推出Grok。在中國，大語言模型競爭更為激烈，截至2024年1月，國產大模型超過二百個。

朱嘉明 經濟學博士、教授，曾任職於聯合國工業發展組織（UNIDO），先後任教於維也納大學和台灣大學，現任橫琴數鏈數字金融研究院學術與技術委員會主席。