AI歷史的重要里程碑
——圖靈機誕生九十周年與達特茅斯會議七十周年

● 朱嘉明

1936與1956年，這兩個看似平凡的年份，發生了人工智能（AI）歷史上最為重要的兩個事件，在人類文明史上鐫刻下不可磨滅的印記。前者見證了年輕的英國數學家圖靈（Alan Turing）提出圖靈機理論，為現代計算奠定了理論基石；後者則誕生了「人工智能」（artificial intelligence）這一術語，開啟了人類對機器智能的系統性探索。2026年，當我們站在這兩個里程碑事件的九十周年和七十周年節點回望歷史，其意義愈發深遠：我們正處於人工智能發展的關鍵轉折點，通用人工智能（artificial general intelligence, AGI）的實現被廣泛預測將在未來數年內到來。站在「技術奇點」的邊緣，回望歷史的火種，既是對先驅的致敬，也是對未來的審視。

一　1936年：圖靈與計算的理論基石

（一）圖靈論文的革命性貢獻

二十世紀30年代，數學界正經歷一場深刻的基礎危機。德國數學家希爾伯特（David Hilbert）提出的數學基礎綱領，試圖將所有數學建立在完備且一致的公理系統之上，其中「判定性問題」（Entscheidungsproblem）成為核心挑戰：是否存在一個算法，能夠判定任何數學命題的真假？這個問題不僅關乎數學的完備性，更觸及了「計算」本質的哲學思考。1931年，哥德爾（Kurt F. Gödel）的「不完備定理」（Incompleteness Theorems）已經動搖了希爾伯特綱領的根基，但對「可計算性」的精確定義仍然懸而未決。

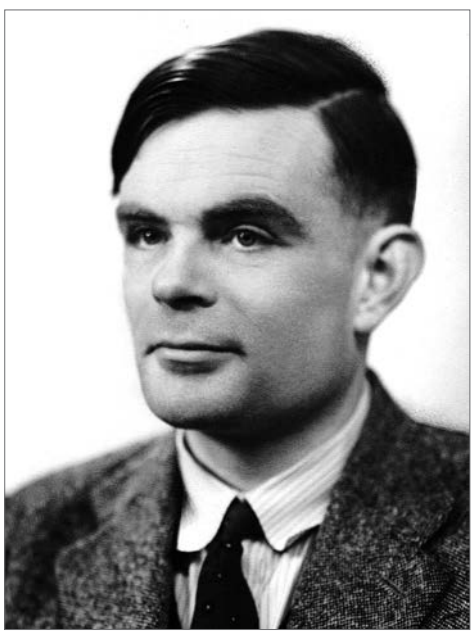
在這樣的時代背景下，年僅二十四歲的圖靈登上了歷史舞台。1936年5月28日，圖靈向倫敦數學學會（London Mathematical Society）提交了題為

〈論可計算數及其在判定性問題上的應用〉(“On Computable Numbers, with an Application to the Entscheidungsproblem”)的論文，該論文於11月12日正式宣讀①。九十年之後再看這篇當時人們不能意識到其深遠意義的文章，可以毫不誇張地說，這也是普羅米修斯(Prometheus)的思想之火。在這篇開創性論文的開頭，圖靈給出了「可計算數」的定義：「『可計算數』可以簡要描述為那些其十進制表達可以通過有限手段計算出來的實數。」他寫道：「根據我的定義，如果一個數的十進制表達可以被機器寫下來，那麼這個數就是可計算的。」

圖靈的天才之處在於，他沒有試圖直接回答希爾伯特的判定性問題，而是退後一步，首先定義了「可計算性」的精確含義。他構想了一個極其簡潔的抽象機器模型：一條無限長的紙帶被劃分為若干方格，每個方格可以寫入符號；機器能夠掃描紙帶上的特定方格，並根據當前狀態和掃描到的符號，按照預設的規則改寫符號、沿紙帶移動或改變自身狀態。圖靈在論文中詳細描述了這一模型：「我們可以把正在計算實數的人比作一台只能具有有限個狀態$q_1, q_2, ..., q_r$的機器，這些狀態將被稱為『m-配置』(m-configurations)。這台機器配備有一條『紙帶』(相當於紙)，紙帶穿過機器並被分成若干段(稱為『方格』)，每個方格可以承載一個『符號』。」他進一步解釋：「機器在任何時刻的可能行為由m-配置q_n和被掃描的符號決定。」

更重要的是，圖靈提出了「通用圖靈機」(Universal Turing Machine)的概念：一個可以模擬任何其他圖靈機的抽象計算模型。他在論文第六節「通用計算機器」(The Universal Computing Machine)寫道：「可以發明一台單一的機器，它可以被用來計算任何可計算序列。如果向這台機器$\mathscr{U}$提供一條紙帶，紙帶的開頭寫有某台計算機器$\mathscr{M}$的標準描述，那麼$\mathscr{U}$將計算出與$\mathscr{M}$相同的序列。」這一洞見本質上預見了現代通用計算機的核心思想：硬件可以保持不變，通過改變軟件(程式)來執行不同的任務。

論者認為，「圖靈的構想與其他人的區別在於，他是通過一台抽象機器來表示的。其特性的驚人之處在於他表明了只要有無限的數據存儲能力，就算用非常有限的基礎操作也足夠實現所有算法。進而，他展示了無論是甚麼算法，都可以設計出一台稱為『通用機器』的機器來實現。這些遠見在現代多用途計算機的發展中扮演着至關重要的角色」②。或者可以說，圖靈機理論的革命性還在於它明確劃定了可計算性的邊界。通過構造「停機問題」，即「是否存在一個通用算法，



圖靈提出了「通用圖靈機」的概念(資料圖片)

能判定任意程式在給定輸入下最終停止運行（而非陷入無窮迴圈）」，圖靈在論文第八節「對角線過程的應用」（Application of the Diagonal Process）證明：「如果存在這樣一個通用過程，那麼就會有一台機器來計算β……這導致矛盾。」這證明了某些問題是任何算法都無法解決的，為計算理論設定了根本性的限制。

值得注意的是，幾乎在同一時期，美國數學家邱奇（Alonzo Church）用完全不同的方法——λ演算（Lambda calculus），獨立地解決了判定性問題。圖靈在論文中提到，邱奇的「『有效可計算性』概念與我的『可計算性』等價，但定義方式非常不同」。兩種完全獨立的形式化系統最終被證明等價，這一事實極大地增強了理論的可信度。這促成了後來被稱為「邱奇—圖靈論題」（Church-Turing thesis）的基本原理：所有直覺上可計算的函數都屬於圖靈機可計算的範疇。這一論題雖然無法被嚴格證明（因為「直覺上可計算」無法被形式化定義），卻成為計算理論和計算機科學的基石，至今沒有反例被發現。荷蘭計算機先驅范德波爾（Willem van der Poel）是少數真正在早期計算機設計中應用圖靈1936年「通用機器」概念的人。事實上，在計算機科學早期發展中，許多先驅在不了解圖靈工作的情況下也取得了偉大的成就③。這說明偉大思想的傳播和被理解需要時間，但最終它們的價值會被歷史充分認可。

（二）為現代計算機科學奠定理論基礎

圖靈在論文第九節對「可計算數的範圍」（The Extent of the Computable Numbers）進行了深入論述。他坦誠地寫道：「試圖證明『可計算數』包含所有自然被認為是可計算的數的所有論證，從根本上說都是對直覺的訴求，因此在數學上相當不令人滿意。」他進而提供了一個令人信服的論證：「計算通常是通過在紙上寫上某些符號來完成的……我假設計算是在一維紙上進行的，即在被分成方格的紙帶上進行的」，並進一步指出：「計算者在任何時刻的行為由他正在觀察的符號和他當時的『心智狀態』決定……我們還假設需要考慮的心智狀態的數量是有限的。」通過這種分析，圖靈論證了他的機器模型可以捕捉所有人類可執行的計算過程。

圖靈的理論天賦很快在實踐中得到驗證。1939年第二次世界大戰爆發後，圖靈被招募到英國政府密碼學校（Government Code and Cypher School），在布萊切利園（Bletchley Park）從事破譯德軍密碼的工作。他設計並改進了用於破譯恩尼格瑪（Enigma）密碼機的電子機械裝置「炸彈機」（Bombe），還開發了稱為「班布里斯姆斯」（Banburismus）的統計技術。

圖靈的工作為整個現代計算機科學奠定了理論基礎。更深遠的影響是，圖靈機理論蘊含着一個哲學命題：如果智能可以被算法化描述，那麼機器也能實現智能。1950年，圖靈在《心靈》（*Mind*）雜誌上發表了另一篇影響深遠的論文〈計算機與智能〉（“Computing Machinery and Intelligence”），明確提出了「機器能思考嗎？」這一問題④。他在文中寫道：由於「思考」和「機器」這兩個

詞無法清晰定義，我們應該用另一個與之密切相關且以相對明確的詞語表達的問題來替代。這就是著名的「圖靈測試」(Turing Test)：如果一台機器在對話中無法被人類區分出來，那麼就可以認為它具有智能。

在這篇論文中，圖靈對機器學習(machine learning)也有先見之明的論述。他在第七節「學習機器」(Learning Machines)中寫道：「與其試圖編寫一個程式來模擬成人的心智，為甚麼不試圖編寫一個模擬兒童心智的程式呢？」這一洞見預見了現代深度學習(deep learning)的核心思想：通過訓練來獲得智能，而非人工編碼規則。這個思想實驗成為連接計算與智能的橋樑，直接啟發了六年後達特茅斯會議的召開。

圖靈的工作曾經對盟軍在二戰中取勝產生了決定性影響。歷史學家估計，布萊切利園的密碼破譯工作可能將戰爭縮短了兩年以上，挽救了數百萬人的生命⑤。然而，這位為人類文明作出巨大貢獻的天才，卻未能在生前得到應有的尊重。1952年，圖靈因同性戀行為被起訴，當時英國法律將其定為刑事犯罪，他接受了化學閹割治療。1954年6月7日夜裏，年僅四十一歲的圖靈結束了自己的生命，在家中去世。圖靈不會想到兩年之後，在美國召開了受他啟發而誕生的人工智能會議。如果他還活着，是否會被邀請參加這個會議？如果參加，他又會為人工智能的發展帶來怎樣的影響和貢獻？誠如論者所言：「圖靈研究思維，可思維卻是那樣神秘，這位解密大師，最終敗給了自己內心的謎。」⑥

二　1956年：達特茅斯會議與人工智能的命名

(一) 達特茅斯會議的核心討論

自1936年圖靈機理論提出後的二十年間，世界經歷了二戰。在慘烈的戰爭中，電子計算機從理論走向現實。1945年，數學家馮．諾依曼(John von Neumann)提出存儲程式計算機概念和架構。1946年2月，世界上第一台通用計算機“ENIAC”(Electronic Numerical Integrator and Computer)於美國賓夕法尼亞大學公開發布。

1952年夏天，在美國貝爾實驗室(Bell Telephone Laboratories)為美國數學家、電子工程師香農(Claude E. Shannon)工作的普林斯頓大學數學系兩位研究生麥卡錫(John McCarthy)和明斯基(Marvin L. Minsky)，受香農影響開始將研究興趣轉移到圖靈機以及是否可用圖靈機作為智能活動的理論基礎⑦。

1955年初，麥卡錫向洛克菲勒基金會(Rockefeller Foundation)申請經費，計劃在達特茅斯舉辦一個夏季研討會。這次會議使得「人工智能」這一術語得以首次正式使用。6月，麥卡錫和香農與基金會生物醫學研究主任莫里森(Robert S. Morison)會面討論這一想法和可能的經費支持。8月31日，麥卡錫、

明斯基、國際商業機器公司(IBM)的羅切斯特(Nathaniel Rochester)及香農聯合簽署了〈達特茅斯人工智能夏季研究項目提案〉("A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence",以下簡稱「達特茅斯提案」)。這份歷史性文件的開頭寫道:「我們建議在1956年夏天於新罕布什爾州(New Hamsphire)漢諾威鎮(Hanover)的達特茅斯學院(Dartmouth College)進行為期兩個月、十人參加的人工智能研究。此項研究在這樣的猜想基礎上進行,即學習以及智能的任何其他特性的每一方面在原則上都能被精確描述,以致可使一台機器來模擬它。」達特茅斯提案繼續闡述:「我們會嘗試尋求如何讓機器使用語言,形成抽象和概念,解決現在留待人類解決的問題,並提升自己。我們認為,如果一群精心挑選的科學家在一起工作一個夏天,就可以在這些問題中的一個或多個方面取得重大進展。」⑧應該注意到,1955年的麥卡錫和明斯基只有二十八歲,羅切斯特三十六歲,香農不過三十九歲,可見年輕科學家對於前沿探索的重要貢獻。

達特茅斯會議於1956年6月18日開幕,歷時八周。這場會議的實際參加者的規模遠超預期。麥卡錫後來列出了多達四十七人的名單,包括多次訪問或短期參與的研究者。核心參與者包括索羅門諾夫(Ray Solomonoff)、紐厄爾(Allen Newell)、西蒙(Herbert A. Simon)、塞繆爾(Arthur L. Samuel)、塞爾弗里奇(Oliver G. Selfridge)等人,他們中的許多人後來成為人工智能領域的奠基人物⑨。

參與者在不同時間到達達特茅斯會議會場,多數只停留較短時間。索羅門諾夫和明斯基住在教授公寓,大多數人則住在漢諾威旅館(Hanover Inn)。這種鬆散而靈活的組織方式,反而營造了一種自由探討的學術氛圍,讓不同



達特茅斯會議參與者(圖片來源:The Minsky Family)

背景的研究者可以深入交流各自的想法。因為此次會議，這所創建於1769年的常春藤大學成為了人工智能策源地。

達特茅斯會議沒有正式的演講議程，它採用完全開放的腦力激盪（brainstorming）形式——參與者聚集在達特茅斯學院數學系頂樓的主教室，各人輪流主導討論，分享自己的研究方向。這種非正式的氛圍被認為極具創造力——儘管由於各人觀點迥異、湊不成統一方向，會議並未產出集體性的研究成果。以下是參與者的發言與核心討論：

第一，紐厄爾與西蒙的「邏輯理論家」（Logic Theorist）程式。會議上令人印象最深的報告來自紐厄爾和西蒙：他們展示了「邏輯理論家」程式，這是第一個能自動證明羅素（Bertrand Russell）與懷特海（Alfred N. Whitehead）合著的《數學原理》（*Principia Mathematica*）中命題邏輯定理的人工智能程式⑩。這篇論文早先投到《符號邏輯雜誌》（*The Journal of Symbolic Logic*）但遭到退稿，「編輯完全沒有理解其中的深意，以『在《數學原理》這個過時的系統中證明一個定理並無成就可言』為由，拒絕了這篇論文」⑪。西蒙曾將程式所證明的一條定理展示給羅素本人——該證明甚至比羅素與懷特海人工推導的版本更為優雅——羅素對此回應頗為欣喜⑫。然而，當時在達特茅斯會議現場，幾乎沒有人意識到這一成果的深遠意義，整體反應相當冷淡。會後，紐厄爾對麥卡錫代表會議作報告一事頗為不滿。達特茅斯會議「最重要的科學成果，或許也是唯一的成果」，正是紐厄爾等人的「邏輯理論家」程式⑬。

第二，明斯基的「幾何定理證明機」（geometry theorem machine）。明斯基在會議期間在紙上設計了一個幾何定理證明機，並提出了以等腰三角形為例的設想，其思路是讓機器透過作圖來產生各子目標的圖形示意，從而引導機器逐步發現並驗證中間步驟所需的輔助定理⑭。羅切斯特在會場與明斯基詳談了這一想法，其後將其介紹給IBM的研究員格蘭特（Herbert L. Gelernter），後者在1960年正式實現並發表了幾何定理證明機的論文⑮。

第三，塞繆爾的跳棋學習程式。塞繆爾講述了他的跳棋程式研究，重點探討「如何引入中間目標」以及「如何對不同類型的棋步評分」。索羅門諾夫在筆記中記錄：「塞繆爾完全感興趣的是學習問題，而不是跳棋本身。」此外，伯恩斯坦（Alex Bernstein）講述國際象棋，探討如何調整權重使機器偏向贏家的走法。

第四，索羅門諾夫的「歸納推理機」（inductive inference machine）。全程參與八周會議的共有三人：索羅門諾夫、麥卡錫和明斯基。索羅門諾夫在會議期間撰寫了175頁筆記，提出了「歸納推理機」的概念，其思路是用符號矩陣表示程式代碼，以此預測未來的符號序列。明斯基後來承認，正是索羅門諾夫關於符號歸納推理的早期研究工作，推動了他放棄神經網絡建模，轉向符號人工智能的研究方向。

第五，阿什比（William R. Ashby）的「穩態調節器」（homeostat）。阿什比主導了一次關於穩態調節器的討論——這是一種能適應環境變化的機電裝置。

他強調，一台構造簡單的機器，若將其內部機制遮蔽，從心理學角度看會顯得極為神奇。麥卡錫對此呼應，認為唯一真正的問題是搜索問題——亦即如何提升搜索效率與速度。索羅門諾夫則在筆記中指出該機器的兩個缺陷：一是沒有對先前解法的記憶，二是求解時間隨機器組件數量呈指數增長。

索羅門諾夫在事後對達特茅斯會議的評價相當嚴苛：「這個研究項目沒甚麼啟發性。收穫最大的是：(1) 寫了報告；(2) 認識了一些有趣的人；(3) 發現這個領域大多數人的思維水平之低；以及 (4) 一些想法：搜索問題可能很重要……」這也提醒我們：歷史里程碑的意義，往往是在後來者的回望中才得以彰顯的。

(二) 會議中形成的三條發展路線

值得留意的是，達特茅斯會議雖然未能形成統一的研究綱領，卻在自由討論中孕育了人工智能此後七十年發展的三條根本路線——符號主義 (symbolism)、聯結主義 (connectionism) 與行為主義 (actionism)。

第一，符號主義路線的確立。以麥卡錫、紐厄爾和西蒙為代表，符號主義認為智能的本質是對符號的操作與推理，即思維就是符號運算。紐厄爾與西蒙在會議上展示的「邏輯理論家」程式是這一路線的第一個具體成果，直接體現了「用符號規則模擬人類推理」的核心理念。會議結束後，兩人沿着這一方向持續深耕，最終在 1976 年發表的論文〈作為經驗探索的計算機科學：符號與搜索〉(“Computer Science as Empirical Inquiry: Symbols and Search”) 中，正式提出「物理符號系統假説」(Physical Symbol System Hypothesis)：智能行為的充分和必要條件是一個物理符號系統的運作 (下詳) ⑯。符號主義此後長期主導人工智能研究，催生了啟發式算法、專家系統與知識工程，在 1980 年代達到頂峰。

第二，聯結主義路線的萌發與轉向。聯結主義的根源可追溯至 1943 年心理學家麥卡洛克 (Warren S. McCulloch) 和數理邏輯學家皮茨 (Walter H. Pitts, Jr.) 建立的神經元模型 (M-P 模型，又稱「麥卡洛克－皮茨模型」或「MCP 模型」)，其核心思路是通過模擬大腦神經元之間的連接機制來實現智能。達特茅斯會議本身見證了這一路線的關鍵轉折。如前所述，索羅門諾夫關於符號歸納推理的研究深刻影響了明斯基，後者在會議期間的深度交流中逐漸形成了轉向符號人工智能的研究取向。這一傾向在此後十餘年間持續演化，最終在 1969 年出版的《感知機》(*Perceptrons: An Introduction to Computational Geometry*) 一書中以公開的學術批判形式付諸於世⑰，對聯結主義研究造成了決定性的打擊，導致神經網絡研究幾乎完全停滯長達十餘年。儘管如此，聯結主義路線並未消亡，而是經由 1980 年代反向傳播 (backpropagation) 算法的重新發現，在二十一世紀以深度學習的形式成為當今最主流的人工智能範式。

第三，行為主義路線的雛形。行為主義 (又稱控制論學派) 的哲學基礎是「感知—行動」循環：智能不需要內部符號表徵，而是在與環境的持續交互與

反饋中湧現。阿什比關於穩態調節器的討論正是這一路線的早期代表——這種能夠自適應環境變化的機電裝置，體現了無內部模型、依靠反饋調節實現適應行為的核心理念。塞繆爾的跳棋程式也蘊含行為主義的元素：程式並不依賴對棋局的完整符號化描述，而是在棋手的博弈中通過評分機制不斷調整策略。

從認知機制角度看，上述三條路線的本質差異在於：符號主義研究抽象推理，注重邏輯可解釋性；聯結主義研究形象感知，依賴數據驅動的自動學習；行為主義研究「感知—行動」，強調與環境的具身交互。達特茅斯會議上，這三條路線的代表人物共聚一堂，卻因視角迥異而未能融合。這既是會議沒有產生統一成果的原因，也恰恰是其歷史價值所在：它為人工智能的多元發展埋下了種子。索羅門諾夫在事後的記錄中也指出了這一隱性分歧的走向：會議促成了一個微妙的轉變，從神經網絡和數值方法逐漸偏向符號方法，從追求通用學習系統轉向解決特定問題的系統，同時也初步顯現了演繹邏輯系統與歸納概率系統之間的路線張力。而這一張力，在此後數十年的人工智能發展歷程中將反覆浮現。

進入二十一世紀尤其是2020年代以來，三條路線的消長格局已發生根本性重組，呈現出明顯的非均衡態勢：

聯結主義憑藉深度學習的崛起，已全面確立其主導地位。以Transformer架構為核心的大語言模型（Large Language Model, LLM），如GPT、Gemini、Claude等，在語言理解、代碼生成、多模態推理等任務上不斷突破人類水平。2025年以DeepSeek-R1於1月開源發布、OpenAI o3系列持續驗證推理擴展律為核心標誌的關鍵轉折在於：大語言模型的智能提升路徑發生根本轉變，不再僅依賴訓練階段的參數擴展，而是更多地通過推理方法優化——即基於強化學習的「思考」過程來實現能力的躍升。這標誌着聯結主義與行為主義路線的深度融合已初步成形。

符號主義雖不再居於主流，卻以新的形式實現了再生。在大語言模型可解釋性需求的推動下，神經符號人工智能（Neuro-Symbolic AI）作為融合方向快速興起：通過將神經網絡的感知學習能力與符號系統的邏輯推理能力相結合，試圖同時解決深度學習「不可解釋」與符號人工智能「知識獲取瓶頸」兩大痼疾。在醫療、金融等對可解釋性有剛性要求的垂直領域，專家系統與知識工程的思路仍不可替代。

行為主義則以強化學習為核心載體持續演進。從AlphaGo到具身智能機器人，強化學習所體現的「感知—行動—反饋」循環正在從遊戲環境向真實物理世界延伸。尤其值得關注的是，強化學習從人類反饋（reinforcement learning with human feedback, RLHF）發展成為訓練大語言模型使其符合人類意圖的關鍵技術，這使得行為主義的核心機制深度嵌入於當代最主流的人工智能訓練範式之中。

三條路線由此完成了一次歷史性的融合：聯結主義提供感知與生成的基礎能力，符號主義提供推理與可解釋性的約束框架，行為主義提供與環境交

互及目標對齊的優化機制。這一融合趨勢，或許正是達特茅斯會議埋下的多元發展種子在七十年後所結出的果實。

(三) 從輝煌到寒冬：人工智能發展的曲折道路

達特茅斯會議的參與者中，有多人後來榮獲圖靈獎（A. M. Turing Award）及其他頂級獎項，堪稱計算機科學史上最密集的獲獎群體之一。圖靈獎被譽為「計算機科學的諾貝爾獎」，達特茅斯會議先後貢獻了四位得主：明斯基（1969）、麥卡錫（1971）、紐厄爾與西蒙（1975年共同獲獎）。四位圖靈獎得主恰好涵蓋了人工智能早期最核心的兩條研究路線：麥卡錫與明斯基代表符號人工智能與知識表示，即以符號和邏輯規則來編碼人類知識，使機器能夠進行形式化推理；紐厄爾與西蒙代表認知模擬與問題求解，即將人類解題的思維過程轉化為可執行的程式。他們的獲獎年份集中於1969至1975年，某種程度上也是對達特茅斯會議精神的遲來認可。

會議之後的十多年間，研究者在符號推理、問題求解、博弈等領域取得了一系列突破。紐厄爾和西蒙等人隨後開發的「通用問題求解器」（General Problem Solver, GPS）試圖用統一的方法解決各類問題⑱；塞繆爾開發的跳棋程式在1962年擊敗了康涅狄格州的冠軍棋手。這些成就似乎驗證了達特茅斯會議的樂觀預期。

然而，好景不長。到1960年代末，人工智能研究遇到了嚴重瓶頸。前文提到1969年明斯基和帕珀特（Seymour Papert）出版了《感知機》一書。這本書詳細分析了單層感知機的數學局限性，證明它無法解決某些簡單問題（如異或問題〔XOR問題〕）。雖然書中只研究單層感知機，並指出多層網絡可能克服這些限制，但單層感知機與多層網絡之間的細微差別被忽略了。該書被廣泛誤解為對整個神經網絡方法的否定，導致聯結主義研究幾乎完全停滯。感知機的發明者羅森布拉特（Frank Rosenblatt）於1971年去世，更為這一研究方向蒙上了陰影。

更大的打擊來自1972年的一份研究報告。英國數學家萊特希爾（James Lighthill）受英國科學研究委員會（Science Research Council）委託，對人工智能研究進行評估。他的報告極為嚴厲，將人工智能研究分為三類：A類（高級自動化）、B類（構建通用智能機器）、C類（計算機模擬神經系統）。萊特希爾認為，B類研究，也就是達特茅斯會議的核心目標，基本上是失敗的，大部分承諾都未能兑現⑲。這份報告的影響是災難性的。英國科學研究委員會大幅削減人工智能研究資金，許多實驗室被迫關閉，研究人員轉向其他領域。這一影響迅速蔓延到美國，引發了第一次「人工智能寒冬」（1974-1980）。大學停止開設人工智能課程，學生避免選擇這一研究方向，風險投資也遠離了人工智能領域。

在這樣的背景下，紐厄爾和西蒙於1976年發表了前文提到的人工智能思想史上的重要文章〈作為經驗探索的計算機科學〉。文章將計算機科學定性為

一門經驗科學，而非純粹的數學或工程學科。他們認為，現實世界中的對象和過程都可以用符號來描述和解釋，而各種問題都可以通過啟發式搜索來獲取答案。也就是說，計算機程式不僅是工具，更是可供觀察和檢驗的科學實驗對象；智能行為的本質是對符號的操縱與搜索過程；科學發現本身也是一種可被程式化的問題求解類型。文章最重要的貢獻是正式提出了「物理符號系統假說」，這一假說奠定了符號主義的理論基礎，認為人類思維與計算機程式在本質上是同構的，都是對符號的處理。

1980年代中期，專家系統的商業應用一度重新點燃了人們對人工智能的熱情。這些基於龐大知識庫和推理機的系統，在醫療診斷、礦物勘探等特定領域展現出商業價值。然而到1987年，專家系統的局限性全面暴露：其知識獲取嚴重依賴人類專家（即「知識獲取瓶頸」），維護成本極其高昂；且系統極為「脆弱」——一旦超出預設的狹窄知識邊界，系統便會徹底崩潰。同時，蘋果公司（Apple Inc.）和IBM的個人電腦崛起，也讓昂貴的專用Lisp計算機硬件市場土崩瓦解。這一泡沫迅速破裂，引發了第二次「人工智能寒冬」（1987-1993）。

這兩次寒冬給人工智能研究留下了深刻教訓：過度承諾和未能兑現會導致嚴重的信任危機。在當今對AGI的討論中，這段歷史仍然具有警示意義——技術突破需要扎實的研究積累，而非空洞的宣傳。

但達特茅斯會議點燃的火種從未真正熄滅。在寒冬中，一些研究者默默堅持。1986年，魯梅爾哈特（David E. Rumelhart）、辛頓（Geoffrey E. Hinton）和威廉姆斯（Ronald J. Williams）重新發現並推廣了反向傳播算法，讓多層神經網絡訓練成為可能[20]。1990年代，隨着統計學習方法興起，人工智能研究從追求人類般的推理轉向從數據中學習。這些轉變並非對達特茅斯會議願景的背離，而是通向同一目標的不同路徑探索。正是在挫折與堅持中，人工智能發展逐漸走向成熟。

三　兩個事件的內在紐帶與技術奇點

圖靈機理論與達特茅斯會議並非孤立的歷史事件，而是智能探索道路上的兩個關鍵節點，構成了從理論到實踐的完美接力。圖靈在1936年的論文中證明了通用計算的可能性，表明任何可以被明確描述的過程都可以被機器執行。如果沒有這一理論基石，達特茅斯提案中關於智能可以由機器來模擬的信念將缺乏堅實的根基。

達特茅斯會議則將圖靈的理論洞見推向工程實踐。從某種意義上說，該會議是對圖靈1950年論文的呼應：既然圖靈提出了判斷機器智能的標準，那麼就應該着手實現。這種從理論到應用的跨越，正是科學進步的典型模式。值得注意的是，1952年麥卡錫和明斯基在貝爾實驗室為香農工作時，香農支持了他們對於圖靈機的研究，清晰地展現了從圖靈到達特茅斯的思想傳承脈絡[21]。

兩個事件的深層聯繫在於它們共享同一個哲學信念：計算主義。前文提到，圖靈在1936年論文的論證隱含地宣稱，思維過程本質上是一種計算過程：「計算者在任何時刻的行為由他正在觀察的符號和他當時的『心智狀態』決定」；達特茅斯提案則明確提出智能是可以由機器模擬的。這兩個論斷共同構成了強人工智能的理論基礎。

這種計算主義哲學具有深刻的顛覆性。它挑戰了人類智能的獨特性，主張心智與身體可以分離，思維可以脫離生物基質在矽基介質上實現。圖靈在1950年的論文中專門討論了這一點，反駁了「神學異議」，即認為思考是不朽靈魂的功能，上帝賦予了人類靈魂但沒有將之賦予動物或機器。無論這一觀點是否完全正確，都極大地拓展了我們對智能本質的理解，並激發了幾代科學家的探索熱情。

圖靈機理論和達特茅斯會議為後世留下了豐厚的理論遺產。圖靈機不僅是計算理論的基礎，也是算法設計、編程語言、操作系統等一切計算技術的源頭。圖靈在論文中建立的「標準描述」(standard description)和「描述數」(description number)系統，實際上已經包含了現代編程和存儲程式計算機的基本思想。達特茅斯會議則開啟了認知科學、神經科學、語言學、哲學等多學科交叉的新領域，推動了人類對自身智能的科學認識。達特茅斯提案中列出的七個研究方向，即自動計算機、語言使用、神經網絡、計算規模理論、自我改進、抽象能力、隨機性和創造性，至今仍然是人工智能研究的核心議題。事實上，當我們審視今天的人工智能研究前沿，會驚訝地發現這七個方向幾乎涵蓋了所有主要領域。

更重要的是，這兩個事件激勵了一代代研究者的精神火種。當人工智能研究在寒冬中徘徊時，正是對達特茅斯會議願景的信念支撐着研究者繼續前行；當算法遇到理論瓶頸時，回到圖靈機的基本原理往往能帶來新的啟發。歷史的價值不僅在於它曾經發生，更在於它持續照亮未來的道路。

九十年後的今天，圖靈機已經從抽象的數學模型演化為無處不在的現實。全球數十億台計算設備，從智能手機到超級計算機，本質上都是通用圖靈機的物理實現。每一次程式運行，都是對圖靈「狀態—符號—規則」框架的演繹；每一個操作系統，都體現着通用圖靈機「一機模擬萬機」的思想。

在人工智能領域，圖靈機理論的意義更加深遠。現代人工智能晶片，無論是圖形處理器(Graphics Processing Unit, GPU)、張量處理器(Tensor Processing Unit, TPU)，還是神經網絡處理器(Neural Processing Unit, NPU)，其底層計算架構都遵循着圖靈機的基本原理。大語言模型的訓練和推理過程，本質上是在龐大的參數空間中執行複雜的符號變換，這正是圖靈機計算範式的擴展。可計算性理論也在提醒我們：即使AGI實現，也必然存在某些超出其能力範圍的問題：圖靈九十年前在論文中證明的「停機問題」，至今仍是智能的根本邊界。

圖靈在1950年的論文中對機器學習也有先見之明的論述。他關於編寫一個模擬兒童心智的程式的想法預見了現代深度學習的核心思想：通過訓練

來獲得智能，而非人工編碼規則。今天，神經網絡從隨機初始化開始，通過大量數據的訓練逐步「成長」為強大的智能系統，正是對圖靈這一思想的完美詮釋。

而在達特茅斯會議七十年後，其願景正在以驚人的速度成為現實。2012年AlexNet深度神經網絡在ImageNet大規模圖像識別競賽中以壓倒性優勢奪冠，2016年AlphaGo戰勝人類圍棋冠軍，2020年代以GPT-3為開端的大語言模型展現出智能湧現，這些里程碑事件一步步驗證着早期人工智能研究者的信念。讓我們逐一審視達特茅斯提案中的七個研究方向在今天的實現情況：

(1) 自動計算機：現代計算機已成為實現智能的物質基礎。

(2) 語言使用與學習：大語言模型如GPT、Claude已經能夠流暢使用自然語言，進行對話、寫作、翻譯等複雜任務；深度學習、強化學習、自監督學習等方法也正在深刻揭示智能學習的本質機制。

(3) 神經網絡：深度學習證明了多層神經網絡確實能夠從數據中自動形成抽象概念。

(4) 計算規模理論：複雜性理論和算法分析已經發展成熟，P vs NP問題成為計算機科學的核心挑戰㉒。

(5) 自我改進：強化學習和元學習展現了機器自我優化的能力，這正是通向AGI的關鍵路徑。

(6) 抽象能力：現代人工智能能夠進行類比推理、遷移學習，展現出抽象思維的萌芽。

(7) 隨機性和創造性：生成式人工智能(generative AI)在圖像、音樂、文本創作中展現出驚人的創造力。

更令人震撼的是實現AGI時間線的急劇壓縮。根據一項對人工智能研究者的大規模調查，在分析了9,800個預測後發現，50%的專家認為AGI將在2040至2050年間實現。而企業界的預測更為激進：馬斯克(Elon Musk)預測2026年將達到AGI，2030年人工智能將超越全人類智能的總和。當前的技術進展似乎支持這些預測：最先進的人工智能系統已經在編程、寫作、數學推理等領域接近甚至超越人類精英水平㉓。

「技術奇點」(technological singularity)——這個由馮．諾依曼提出、科幻作家文奇(Vernor Vinge)推廣的概念，如今已是可觸及的未來㉔。奇點理論認為，當人工智能的能力超越人類並能夠自我改進時，技術進步將進入失控的指數增長階段，人類文明將發生根本性的、不可逆轉的變革㉕。

發明家庫茲韋爾(Ray Kurzweil)基於「加速回報定律」(The Law of Accelerating Returns)預測，奇點將在2045年到來。他的計算表明，技術進步並非線性而是指數級的，每一次突破都為下一次突破提供更強大的工具㉖。從這個角度看，圖靈機理論和達特茅斯會議正是奇點到來的必要準備。前者確立了通用計算的理論可能性，後者開啟了智能工程化的實踐探索，二者共同為通向奇點的未來鋪路。

關於奇點的形態存在不同觀點。庫茲韋爾傾向於「漸進攀升」的觀點，認為技術奇點將是一個逐步演化的過程㉗。而文奇則強調「快速自我改進」的可能性，認為一旦人工智能獲得自我優化能力，進展速度將呈指數級爆炸㉘。還有一些專家認為，奇點可能像電力一樣悄然到來，不平衡地擴散：某些領域率先實現超人類智能，而其他領域仍需時日㉙。

馬斯克的預測更具衝擊力：如果2026年實現AGI，那麼從AGI到超人類智能的跨越可能只需幾年時間。專家調查顯示，AGI達成後二三十年內可能發展至超級智能㉚。一旦人工智能獲得自我改進的能力（這正是達特茅斯提案中列出的一個研究方向），「智能爆炸」將以人類難以想像的速度展開。這意味着，從1936年圖靈機理論提出到技術奇點到來，整個過程可能不超過110年，在人類文明史上不過是轉瞬之間。

四　紀念的意義：站在巨人肩膀上

回望人工智能歷史，最令人驚歎的是其先驅者的遠見卓識。1936年，當圖靈在論文中寫下「根據我的定義，如果一個數的十進制表達可以被機器寫下來，那麼這個數就是可計算的」時，世界上還沒有一台電子計算機；他卻已經預見了通用計算時代的到來，甚至在論文中詳細描述了通用計算機器的工作原理。1956年，當達特茅斯提案的四位作者寫下他們對於機器可以模擬智能的信念時，計算機還是佔據整個房間、需要專人維護的龐然大物；他們卻相信機器終將展現出人類般的智能，甚至在提案中列出了「自我改進」這一今天仍然是AGI核心挑戰的研究方向。

這種先驅遠見並非憑空臆想，而是建立在深刻的理論洞察和大膽的想像力之上。圖靈沒有被當時技術的限制所束縛，而是從第一性原理出發思考計算的本質，他在論文中對人類計算過程進行的細緻分析，展現了非凡的抽象能力。達特茅斯會議的組織者也沒有因為早期技術的原始而氣餒，而是對智能的可實現性保持堅定信念。正是這種超越時代的思考，為後世指明了方向。

更值得致敬的是，先驅者面對質疑和挫折時的勇氣與堅持。圖靈的一生充滿悲劇色彩：他的理論在當時並未被充分理解，戰時的關鍵貢獻因保密法而不為人知，其後被起訴以至早逝更令人扼腕。人工智能研究則經歷了兩次漫長的寒冬，萊特希爾報告的嚴厲批評、感知機爭議帶來的打擊、專家系統泡沫的破裂，都曾讓這一領域瀕臨危機。許多研究者在資金斷絕、前景黯淡時仍然堅守信念，正是這種堅持，讓他們熬過了苦寒的拂曉，收穫了今天的成果。

本文旨在回顧人工智能歷史上兩大事件，給當代帶來重要的啟示：第一，基礎理論的長遠價值。圖靈機理論在提出時看似純粹的數學思辨，與實際應用相距甚遠。然而，正是這樣的理論創新成為整個信息時代的基石。

第二，跨學科合作的重要性。達特茅斯會議匯聚了數學家、工程師、心理學家、語言學家等不同背景的學者，正是這種多元視角的碰撞催生了人工智能這一交叉學科。第三，對未知保持謙卑與敬畏。圖靈在論文中證明了可計算性的邊界，明確指出某些問題無法通過算法解決；達特茅斯會議參與者雖然樂觀，但並沒有宣稱能立即造出完美的智能機器。

歷史對先驅的認可往往姍姍來遲。直到2013年，英國女王才正式赦免了圖靈的同性戀罪。2014年電影《模仿遊戲》(*The Imitation Game*)讓圖靈的故事為大眾所知。這些遲來的認可提醒我們：真正具有革命性的思想，其價值可能需要幾十年甚至更長時間才能被充分認識。在人們紀念圖靈機誕生九十周年、達特茅斯會議召開七十周年的時刻，唯有以堅忍不拔的精神進行永無止境的科學探索，方能應對日益被技術加速主義所主導的世界。

註釋

① 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。下引不再另註。

② 戴維斯(Martin Davis)著，高曉溯譯：〈算法、方程和邏輯〉，載庫珀(S. Barry Cooper)等著，堵丁柱等譯：《永恆的圖靈：20位科學家對圖靈思想的解構與超越》(北京：機械工業出版社，2018)，頁11。

③ Edgar Daylight, "Turing's 1936 Paper and the First Dutch Computers" (19 August 2013), https://cacm.acm.org/blogcacm/turings-1936-paper-and-the-first-dutch-computers/.

④ 參見A. M. Turing, "Computing Machinery and Intelligence", *Mind* LIX, issue 236 (1950): 433-60。下引不再另註。

⑤ Jack Copeland, "Alan Turing: The Codebreaker Who Saved 'Millions of Lives'" (19 June 2012), www.bbc.com/news/technology-18419691.

⑥ 霍奇斯(Andrew Hodges)著，孫天齊譯：《艾倫．圖靈傳：如謎的解謎者》(長沙：湖南科學技術出版社，2012)，頁506。

⑦⑭㉑ John McCarthy, "The Dartmouth Workshop—as Planned and as It Happened" (30 October 2006), www-formal.stanford.edu/jmc/slides/dartmouth/dartmouth/node1.html.

⑧ 參見John McCarthy et al., "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955", *AI Magazine* 27, no. 4 (2006): 12-14。下引不再另註。

⑨ 關於會議的主要內容和發言，參見Grace Solomonoff, "Ray Solomonoff and the Dartmouth Summer Research Project in Artificial Intelligence, 1956", https://raysolomonoff.com/dartmouth/dartray.pdf, 1-28。下引不再另註。

⑩ John McCarthy, "The Dartmouth Workshop—as Planned and as It Happened". 另可參見Herbert A. Simon, *Models of My Life* (New York: Basic Books, 1992), 189。西蒙在該回憶錄中稱，1955和1956年是自己學術生涯最重要的兩年。

⑪ Daniel Crevier, *AI: The Tumultuous History of the Search for Artificial Intelligence* (New York: Basic Books, 1993), 46.

⑫⑬ Pamela McCorduck, *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence* (Natick, MA: AK Peters, 2004), 167; 127-29.

⑮ H. Gelernter et al.,"Empirical Explorations of the Geometry Theorem Machine", conference paper presented at Western Joint IRE-AIEE-ACM Computer Conference (3-5 May 1960), https://dl.acm.org/doi/10.1145/1460361.1460381, 143-49.

⑯ Allen Newell and Herbert A. Simon,"Computer Science as Empirical Inquiry: Symbols and Search", *Communications of the ACM* 19, issue 3 (1976): 113-26.

⑰ Marvin Minsky and Seymour Papert, *Perceptrons: An Introduction to Computational Geometry* (Cambridge, MA: MIT Press, 1969).

⑱ GPS是由紐厄爾、西蒙及肖(J. C. Shaw)於1957年開發的早期人工智能程式。其核心思想是：人類解決問題的思維過程可以被形式化、程式化。GPS採用「手段—目標分析」(means-ends analysis)策略：不斷比較當前狀態與目標狀態之間的差距，並選擇能縮小該差距的操作。它不針對某一特定領域(如下棋或數學)，而是希望建立一套通用的推理框架，適用於邏輯證明、謎題破解、語言理解等多種問題。這體現了早期人工智能研究對「強人工智能」的樂觀願景。GPS雖具開創性，但最終因「組合爆炸」問題(搜索空間隨問題複雜度急劇增大)而在實際應用中受限，促使後來研究者轉向更專門化的專家系統方向。

⑲ James Lighthill,"Artificial Intelligence: A General Survey" (July 1972), www.chilton-computing.org.uk/inf/literature/reports/lighthill_report/p001.htm.

⑳ David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams,"Learning Representations by Back-Propagating Errors", *Nature* 323, no. 6088 (1986): 533-36.

㉒ 其中，P vs NP問題被公認為理論計算機科學的核心未解難題：P類問題係指可於多項式時間(polynomial time)內求解之決策問題，NP類問題則係指其解可於多項式時間內加以驗證之決策問題。

㉓ Cem Dilmegani and Sıla Ermut,"AGI/Singularity: 9,800 Predictions Analyzed" (24 February 2026), https://research.aimultiple.com/artificial-general-intelligence-singularity-timing/.

㉔ S. Ulam,"John von Neumann 1903-1957", *Bulletin of the American Mathematical Society* 64, no. 3 (1958): 1-49; Vernor Vinge,"The Coming Technological Singularity: How to Survive in the Post-Human Era", paper for the VISION-21 Symposium (NASA Lewis Research Center and the Ohio Aerospace Institute, 30-31 March 1993), https://edoras.sdsu.edu/~vinge/misc/singularity.html.

㉕㉖㉗ Ray Kurzweil,"The Law of Accelerating Returns" (1 January 2001), www.writingsbyraykurzweil.com/the-law-of-accelerating-returns.

㉘ Vernor Vinge,"The Coming Technological Singularity".

㉙ Robin Hanson, *The Age of Em: Work, Love, and Life When Robots Rule the Earth* (Oxford: Oxford University Press, 2016).

㉚ Adolphus Upton,"Musk's 2026 AI Prediciton: Timeline to Superhuman Intelligence", https://waterloow.com/2026/01/28/musk-superhuman-ai-prediction-2026/.

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、台灣大學等大學，現任橫琴數鏈數字金融研究院學術與技術委員會主席。