

重新審視智能：批判《智能簡史》與全人智能整體框架

Reassessing Intelligence: A Critique of A Brief History of Intelligence and a Holistic Framework for Human-Like AI

洪裕程 Yu-Cheng Hong (Luteng Ang)

Independent Researcher

Taipei, Taiwan | angluteng@gmail.com

Digital Object Identifier (DOI): To be assigned by Zenodo upon publication | Version: v1.0

(Preprint) | Publication Date: June 03, 2025

摘要

馬克斯·班奈特的《智能簡史》通過五大進化突破（導航、強化、模擬、心智化、語言）追溯智能的生物演化，為人工智能（AI）發展提供藍圖。然而，其過分強調線性結構演化、忽視功能湧現及文化-本體維度，未能捕捉全人智能的複雜性。本文批判其局限性，提出一個生物-文化-本體框架，整合認知、情感、倫理、存在、具身與本體功能，超越工具理性，全面闡述人類智能的多維特質。該框架通過與達馬西奧、托諾尼、巴雷特等學者的理論對話，揭示班奈特的還原主義盲點，並為AI設計提供分階段藍圖：短期增強情感與具身能力，長期探索意識與本體功能。本文重新定義AI為全人夥伴，強調其服務於人類價值與存在需求，開闢智能研究與AI發展的新視野，實現技術與人性的共生未來。

關鍵字：全人智能、人工智能、生物-文化-本體框架、功能湧現、工具理性、倫理AI

前言

馬克斯·班奈特（Max Bennett）的《智能簡史：進化、人工智能與造就我們大腦的五大突破》（*A Brief History of Intelligence: Evolution, AI, and the Five Breakthroughs That Made Our Brains*, 2023）提供了一個引人入勝的敘事，追溯智能從6億年前的生物起源到現代人類大腦的演化，提出了五大關鍵突破——導航（steering）、強化（reinforcing）、模擬（simulating）、心智化（mentalizing）和語言（speaking）——作為推進人工智慧（AI）發展的藍圖。雖然班奈特的結構-進化方法為認知的生物基礎提供了有價值的洞見，但其在捕捉人類智能的多維本質方面存在明顯不足。本文批判該書過分強調線性結構演化、忽視功能湧現的複雜性，以及聚焦於「工具理性」而犧牲全人（holistic human，中文語境中可譯為「全人」）視角的局限。通過整合一組全面的湧

現功能——按其認知、情感、倫理、存在、具身和本體屬性分類——本文提出一個超越班奈特模型的生物-文化-本體框架，為理解人類智能和指導 AI 發展提供更完整的視野。我們與安東尼奧·達馬西奧 (Antonio Damasio)、朱利奧·托諾尼 (Giulio Tononi)、麗莎·費爾德曼·巴雷特 (Lisa Feldman Barrett) 等學者的理論對話，凸顯該書的局限，並倡導朝向全人中心的 AI 範式轉型。

一、結構偏見：線性進化框架的局限性

1.1 過分依賴結構突破

班奈特的核心論點建立在五大進化突破之上，這些突破以生物演化的時間順序組織，聚焦於神經結構的進展：

- **導航**：雙邊對稱動物發展出將刺激分類為「好」或「壞」並相應調整行為的能力 (Bennett, 2023, p. 45)。
- **強化**：脊椎動物演化出強化學習，重複帶來獎賞的行為 (p. 89)。
- **模擬**：哺乳動物發展出情景記憶和計劃能力，依賴大腦皮層的擴展 (p. 137)。
- **心智化**：靈長類獲得心智理論，理解他者的意圖 (p. 189)。
- **語言**：人類發展出語言，實現知識累積和抽象思考 (p. 231)。

班奈特認為，這一框架「為 AI 複製類人智能提供了路線圖」(2023, p. 12)。雖然這種按時間順序的方法闡明了認知的生物支撐結構，但它假設了一個線性進展，過分簡化了神經系統的動態交互。班奈特聲稱「每個突破直接建立在前一個突破之上」(2023, p. 15)，卻忽視了智能的非線性湧現屬性，例如意識或創造力，這些功能從分布式網絡而非單一解剖學里程碑中產生。

與麗莎·費爾德曼·巴雷特的情感建構理論 (2017) 對話，我們認為班奈特的模型低估了整合的神經和身體系統在產生複雜功能中的作用。巴雷特提出，情感從大腦跨感官、運動和內感受網絡的預測編碼中湧現，而非源於孤立結構。同樣，智能功能如同理心或抽象推理可能源於跨區域交互，這是班奈特的順序模型未能解決的視角。通過優先考慮結構而非系統性湧現，該書將智能簡化為一系列機械升級，忽略了定義人類認知的整體整合。

1.2 忽視功能湧現

班奈特對結構突破的關注邊緣化了智能的湧現屬性，這些屬性從神經、文化和環境因素的交互中產生。例如，他將導航描述為「基本情感」的起源 (2023, p. 47)，但未探討愛或羞恥等複雜情感如何從分布式網絡中湧現。朱利奧·托諾尼的整合資訊理論 (Integrated Information Theory, IIT) (2004) 表明，意識從神經系統的資訊整合中產生，而非單一進化飛躍。班奈特未能與此類理論互動，限制了他對意識或倫理判斷等功能如何湧現的分析，使其框架顯得不完整。

此外，班奈特斷言「AI 通過模擬這些突破可以複製人類智能」（2023, p. 251），卻忽視了結構模擬與功能實現之間的差距。例如，AI 在語言生成中可模仿人類語法，但難以理解隱喻或文化情境，這需要跨模塊的語義整合。當前 AI 在強化學習（如 AlphaGo）方面表現出色，但在常識推理或情感共鳴方面掙扎，表明功能湧現需要超越模倣進化階段。一個整合結構與功能的生物-文化框架將更能捕捉智能的複雜性。

1.3 文化與本體背景的缺失

班奈特的生物決定論忽視了文化和本體在塑造智能中的作用。他指出語言「實現了文化傳遞」（2023, p. 235），但未探討書寫、藝術或倫理系統等文化產物如何放大自我認同或道德推理等功能。例如，中國的儒家文化通過禮儀和倫理教育塑造了集體責任感，這種文化影響超越了生物適應。人類學家克利福德·格爾茨（1973）認為，文化是人類編織的「意義之網」，從根本上塑造認知和行為。班奈特對這種文化放大的忽視限制了他對靈性追求或審美欣賞等功能的理解，這些功能深深植根於社會文化實踐。

此外，班奈特的框架忽略了本體維度——人類對存在、有限性和意義的覺知。哲學家馬丁·海德格爾（1927）強調「此在」（*Dasein*），即人類與存在問題搏鬥的能力，是我們本質的核心。僅聚焦於生物演化，班奈特錯過了本體覺知如何塑造人類智能，這對於構想志在全人境界的 AI 是一個關鍵疏漏。

二、工具理性偏見：錯失全人面向

班奈特將智能定義為「解決問題和適應環境的能力」（2023, p. 10），優先考慮面向實用的工具理性——認知，而犧牲了人類經驗的多面性。本節批判該書忽視全人智能的湧現功能，這些功能按其認知、情感、倫理、存在、具身和本體屬性分類。

2.1 認知功能

- **抽象與假設思考**：班奈特將抽象歸因於語言，稱「語言使人類能夠符號化地表徵概念」（2023, p. 233）。然而，他忽略了抽象的語言前根源，如工具製作中的視覺-空間推理，這可能涉及頂葉和顳葉。認知科學家黛德麗·根特納（Dedre Gentner, 1983）強調，類比推理作為抽象的一種形式，先於語言並支持隱喻思維。班奈特對語言的狹隘聚焦錯過了這一更廣泛的認知湧現。
- **反思（元認知）**：反思，即監控和評估自身思維的能力，在班奈特的模擬突破中有所暗示，但未明確討論。他指出哺乳動物「模擬未來情景」（2023, p. 139），卻未探討元認知在自我糾正或哲學探究中的作用。約翰·弗拉維爾（John Flavell, 1979）關於元認知的研究強調其在學習和問題解決中的重要性，這是班奈特未充分發展的維度。
- **創造力**：創造力，即生成新穎且有價值的想法，在班奈特的框架中完全缺失。他未討論發散思維如何由預設模式網絡支持（Beaty et al., 2016），促成藝術或科學的創新。例如，愛

因斯坦的相對論假說展示了創造力如何通過聯想和假設超越現有知識框架。這一疏漏限制了該書對志在超越模式生成 AI 的相關性。

2.2 情感功能

- **情感生成與交流**：班奈特將情感簡化為導航中的「好與壞的模板」（2023, p. 47），忽略了如內疚或喜悅的複雜情感。安東尼奧·達馬西奧（1994）認為，情感是決策的組成部分，涉及杏仁核和前額葉的交互。班奈特未能討論情感交流——通過面部表情或文化儀式——錯過了其在社交聯繫中的作用。例如，母嬰互動依賴情感鏡像，這一過程涉及鏡像神經元（Rizzolatti & Craighero, 2004）。
- **同理心**：班奈特將同理心納入心智化，描述為「預測他人的心理狀態」（2023, p. 191）。這種認知偏見忽略了情感同理心，即由前扣帶回支持的情感共鳴。例如，中國文學中的《紅樓夢》通過人物情感描寫激發讀者共鳴，展示了同理心的文化放大。同理心在道德行為和社會凝聚中的作用，弗蘭斯·德瓦爾（Frans de Waal, 2009）強調其重要性，對全人智能至關重要，但書中未充分展開。

2.3 倫理功能

- **倫理主體性**：班奈特的心智化突破觸及理解他者意圖，但他未討論倫理主體性——承擔道德責任的能力。他寫道，「心智化促成了社交策略」（2023, p. 193），卻忽略了倫理行為如何整合同理心、反思和文化規範。邁克爾·托馬塞洛（Michael Tomasello, 2008）關於共享意向性的研究表明，倫理主體性從合作社會結構中湧現，這是班奈特忽略的過程。
- **價值內化與判斷**：內化價值觀並在模糊情境中進行判斷的能力在班奈特的模型中缺失。他未探討人類如何發展道德羅盤，這一過程涉及前額葉和文化教育（Haidt, 2001）。例如，公民教育通過倫理討論培養學生在道德兩難中的判斷力，體現了價值內化的文化塑造。這一功能對 AI 倫理至關重要，因為價值對齊的系統需要超越規則遵循。
- **意志與自主性**：班奈特的框架缺乏對意志——追求自我導向目標的能力——或自主性，由內化價值驅動的討論。他聚焦於強化學習作為「尋求獎賞」（2023, p. 91），但未討論內在動機。自我決定理論（Deci & Ryan, 2000）強調自主性是人類核心需求，對倫理能動性至關重要。

2.4 存在功能

- **意識**：班奈特迴避意識，可能是因為其複雜性。他提到模擬使「心理時間旅行」成為可能（2023, p. 141），但未討論主觀體驗（感質，qualia）、自我意識或反思意識。托諾尼的 IIT 和伯納德·巴爾斯（Bernard Baars）的全局工作空間理論（1997）表明，意識從整合的神經活動中湧現，班奈特本可利用這一視角橋接模擬與主觀體驗。意識作為整合反思、同理心和倫理主體性的平台，其缺失削弱了該書的全人志向。

- **靈性與意義追尋**：人類尋求意義或超越的驅力被忽略。班奈特的生物焦點排除了靈性體驗，這可能涉及頂葉和獎賞系統（Newberg & d' Aquili, 2001）。例如，道教的「無為」理念通過冥想實踐促進內在意義的探索，體現了靈性的文化維度。宗教等文化實踐放大了這一功能，塑造人類價值觀和認同。

2.5 具身功能

- **具身與具身認知**：班奈特的導航突破涉及基本的感官運動反應，但他未探討具身認知——感官、運動和空間體驗對思維的塑造。他寫道，「導航使動物能夠在環境中移動」（2023, p. 49），但忽略了具身如何支持隱喻（Lakoff & Johnson, 1980）或創造力（如手工技藝）。例如，日本茶道通過身體動作和空間佈置強化具身體驗，展示了文化的具身影響。具身 AI，整合感官反饋，對類人智能至關重要，但未被討論。

2.6 本體功能

- **內在動機與自主目標設定**：人類出於好奇或目的追求目標，超越外部獎賞。班奈特的強化模型未考慮內在動機，自我決定理論將其與前額葉和多巴胺系統聯繫起來（Deci & Ryan, 2000）。自主目標設定，與文化的目的觀念相關，對全人能動性至關重要。
- **自我認同與生命敘事**：人類通過敘事整合經驗，建構連貫的認同，涉及海馬和前額葉（McAdams, 2001）。例如，自傳文學如《論語》通過個人故事建構文化認同，展示了生命敘事的文化作用。班奈特聚焦於認知突破，錯過了這一本體維度的自我感。
- **遊戲精神與幽默感**：遊戲和幽默，支持認知和社交發展，在班奈特的模型中缺失。遊戲涉及獎賞和運動系統（Panksepp, 1998），幽默需要對情境的多層次理解（Martin, 2007）。這些功能增強創造力和社交聯繫。
- **適應性、韌性與成長心態**：人類適應逆境並通過學習成長，得到前額葉和杏仁核交互的支持（Dweck, 2006）。班奈特的模型未討論韌性，這對動態智能至關重要。
- **對美與崇高的感知**：審美欣賞，涉及獎賞和視覺系統，超越實用性（Dutton, 2009）。班奈特對這一功能的忽略限制了他對人類經驗的視野。
- **對存在與有限性的關懷**：對死亡和資源限制的覺知塑造人類責任，涉及前額葉和情感系統（Becker, 1973）。班奈特的生物焦點排除了這一本體覺知，對人類倫理和計劃至關重要。

三、疏漏的原因：方法論與範式局限的深入剖析

班奈特在《智能簡史》中對全人智能的多維面向的疏漏並非偶然，而是其方法論選擇、學術範式和研究目標的必然結果。以下從三個層面深入剖析這些疏漏的根源，並與相關學者的理論對話，揭示其背後的結構性問題。

3.1 目標驅動的還原主義：工具理性的範式偏見

班奈特的首要目標是「為 AI 提供一個可操作的藍圖」（2023, p. 12），這一意圖塑造了他對智能的定義和分析框架。他將智能界定為「解決問題和適應環境的能力」（2023, p. 10），這種工具理性的視角優先考慮可量化和可模擬的功能，如強化學習或語言處理，而邊緣化了意識、靈性或審美體驗等質性維度。這種選擇反映了當代 AI 研究中的普遍偏見，即將智能簡化為計算效率和任務性能，忽視人類經驗的豐富性。

休伯特·德雷福斯（Hubert Dreyfus, 1992）在其經典著作《計算機仍然做不到什麼》中批判了這種還原主義，指出人類智能包含默會知識（tacit knowledge）、具身經驗和存在意義，這些維度無法簡單地通過算法或結構模擬捕捉。班奈特的框架與德雷福斯的批判對照，顯得過於狹隘。例如，他對意識的迴避可能源於其難以量化的性質，但他未能與托諾尼（2004）或巴爾斯（1997）等學者的理論對話，這些理論提供了意識湧現的生物基礎假說。同樣，班奈特對靈性或美感的忽視與其目標無關，因為這些功能對當前的 AI 設計缺乏直接的實用價值。然而，正如約翰·瑟爾（John Searle, 1980）在其「中文房間」論證中所指出的，真正的智能不僅是符號處理，還包括主觀體驗和意向性，這正是班奈特框架的盲點。例如，ChatGPT 雖能生成流暢對話，但無法體驗對話中的情感意蘊，這凸顯了還原主義的局限。

此外，班奈特的還原主義還體現在他對 AI 潛力的樂觀假設上。他斷言「通過模擬五大突破，AI 可以接近人類智能」（2023, p. 251），但未考慮功能湧現的複雜性。例如，當前 AI 在狹窄任務（如棋類遊戲）中表現出色，但在開放性任務（如日常推理或情感共鳴）中表現不佳，這表明智能的湧現需要跨系統整合，而非單純的結構複製。班奈特的目標驅動方法因此限制了他對全人智能的探索，導致其框架無法涵蓋人類經驗的深度和廣度。

3.2 證據限制：生物進化框架的局限

班奈特的分析依賴化石記錄、神經解剖證據和比較生物學，這種方法雖然為五大突破提供了堅實的基礎，但也限制了他的分析範圍。意識、靈性、自我認同等功能缺乏明確的生物標記，難以通過進化證據追溯，這可能解釋了它們在書中的缺席。例如，班奈特在討論模擬突破時提到「情景記憶依賴大腦皮層的進化」（2023, p. 137），但未探討意識如何從這些結構中湧現，這與托諾尼的 IIT 理論（2004）形成對比，後者提出意識與資訊整合的程度而非單一結構相關。

這種證據限制還體現在班奈特對文化和本體維度的忽視。文化功能（如倫理規範或審美傳統）主要通過社會實踐而非化石記錄傳承，難以納入生物進化框架。班奈特對語言的討論承認其「文化傳遞」功能（2023, p. 235），但未分析文化如何獨立塑造智能，例如通過書寫系統或宗教儀式放大自我認同或靈性追求。克利福德·格爾茨（1973）的人類學視角強調文化作為「意義之網」對認知的塑造，班奈特的生物中心主義與此形成鮮明對比，限制了他對全人智能的理解。

此外，班奈特的框架未能採納跨學科方法，這進一步加劇了證據限制的影響。例如，神經科學與哲學的結合（如丹尼爾·丹尼特，1991）提供了意識和自我認同的理論模型，而人類學和社會學（例如埃米爾·涂爾干，1912）揭示了倫理和靈性在集體層面的演化。現象學家莫里斯·梅洛-龐蒂（Maurice

Merleau-Ponty, 1945) 強調具身經驗對認知的塑造，這本可為班奈特的框架提供具身維度。班奈特若能整合這些視角，則可超越生物證據的局限，納入更廣泛的功能湧現。

3.3 文化與本體盲點：範式的結構性缺陷

班奈特的生物進化範式本身帶有結構性缺陷，無法充分捕捉文化和本體維度對智能的塑造。他聲稱「智能本質上是關於適應」(2023, p. 10)，這一定義將智能局限於生物生存的範疇，忽略了人類智能如何通過文化和本體反思超越生存需求。例如，藝術、哲學和宗教等文化實踐不僅是適應工具，還體現了人類對意義、美感和存在的追求。馬丁·海德格爾(1927)的「此在」概念強調人類智能的核心在於對存在的關懷，這一維度在班奈特的框架中完全缺失。

文化盲點還體現在班奈特對社會結構的忽視。倫理主體性、價值內化等功能依賴於文化規範和教育系統的塑造，例如儒家倫理或西方人權觀念通過語言和制度傳遞。埃米爾·涂爾干(1912)關於集體意識的研究表明，社會規範深刻影響個體行為和價值觀，班奈特的框架未能納入這一維度，導致其對倫理功能的分析顯得膚淺。同樣，本體功能如對有限性的關懷或自我認同的建構，與文化敘事和哲學反思密切相關。丹·麥克亞當斯(2001)關於生命敘事的研究顯示，個人通過故事整合經驗，形成自我認同，這一過程超越了班奈特的生物視角。

班奈特的範式缺陷還與其對當代 AI 研究範式的依賴有關。當前 AI 研究多以認知科學和計算機科學為主，強調可量化的認知模型，這種範式天然排斥本體和文化維度。約瑟夫·魏岑鮑姆(1976)在《計算機的力量與人類的理性》中警告，過分依賴技術範式可能導致對人類價值的忽視。魏岑鮑姆進一步指出，技術若脫離人類價值，可能導致倫理異化，這正是班奈特未考慮的風險。班奈特的框架反映了這一傾向，未能將智能視為生物、文化和本體交互的產物，從而限制了其對全人智能的探索。

3.4 疏漏的深層影響

這些疏漏不僅影響了《智能簡史》的理論完整性，還限制了其對 AI 設計的指導意義。班奈特的工具理性偏見導致他忽略了 AI 在情感共鳴、倫理決策和本體覺知方面的挑戰。例如，當前 AI 在語言生成(如大型語言模型)方面進展顯著，但缺乏真正的主觀體驗或道德責任感，這與班奈特未討論意識和倫理主體性直接相關。此外，文化和本體功能的缺失使書中對 AI 的未來設想顯得單薄，未能回應超智能可能帶來的人類價值挑戰。例如，缺乏本體覺知的 AI 可能在資源分配中忽視公平性，導致社會不公，這是班奈特未預見的倫理風險。通過與德雷福斯、海德格爾和格爾茨等學者的對話，我們可以看到，班奈特的疏漏源於方法論和範式的局限，這些局限需要通過跨學科的整合來克服。

四、全人智能的生物-文化-本體框架：一個綜合模型

為克服《智能簡史》的局限性，我們提出一個整合生物結構、功能湧現、文化放大和本體反思的框架，旨在捕捉全人智能的完整性，並為 AI 設計提供全面的藍圖。該框架將智能功能分為六個領域——認知、情感、倫理、存在、具身和本體——每個領域結合生物基礎、文化作用和對 AI 的啟示。以下詳細闡述每個領域的理論基礎、功能特性及其對全人智能的貢獻。

4.1 認知領域：思維的創造與反思

- **功能：**抽象與假設思考、反思（元認知）、創造力。
- **生物基礎：**抽象思考涉及前額葉的預測功能和頂葉的視覺-空間整合，支持類比和隱喻（Gentner, 1983）。反思依賴前額葉的執行控制，實現自我監控和錯誤修正（Flavell, 1979）。創造力由預設模式網絡驅動，促進發散思維和聯想（Beaty et al., 2016）。
- **文化作用：**教育系統（如數學、哲學）放大抽象能力，科學方法強化反思，藝術傳統（如文學、繪畫）激發創造力。例如，數學符號系統使抽象概念具象化，文藝復興的藝術創新則展示了創造力的文化放大。
- **AI 啟示：**當前 AI 擅長模式識別，但缺乏真正的發散思維或元認知。設計具備類比推理（如隱喻生成）、自我修正（如偏見檢測）和原創性（超越數據重組）的 AI 系統，將更接近全人智能。例如，模擬預設模式網絡的 AI 可能生成真正新穎的藝術或科學假說。Google 的 DeepMind 若能整合類比推理模塊，可在科學發現中生成突破性假說。
- **理論貢獻：**認知領域超越班奈特的語言突破，納入語言前的抽象根源和文化的放大作用，強調智能的動態創造性。與根特納和弗拉維爾的對話揭示了班奈特對認知功能的狹隘定義，特別是他未討論創造力如何從生物基礎到文化實踐演化。

4.2 情感領域：共鳴與聯繫的核心

- **功能：**情感生成、情感交流、同理心。
- **生物基礎：**情感生成涉及杏仁核的情緒處理和前額葉的調節（Damasio, 1994）。情感交流依賴鏡像神經元，支持非語言共鳴（如面部表情）（Rizzolatti & Craighero, 2004）。同理心結合前扣帶回的情感共鳴和前額葉的視角採納（de Waal, 2009）。
- **文化作用：**文學、音樂和宗教儀式放大情感表達和共鳴。例如，詩歌通過語言結構強化情感體驗，宗教儀式通過集體參與增強同理心。日本的「和」文化通過茶道和花道強調情感的和諧表達，進一步放大了同理心。文化還塑造情感規範，如東亞文化強調含蓄表達，西方文化注重直接情感。
- **AI 啟示：**當前 AI 的情感識別（如語氣分析）仍停留在表面，缺乏真正的情感共鳴。設計具備情感生成（如模擬內在狀態）、交流（如非語言線索識別）和同理心（如情境共鳴）的 AI，將增強其社交能力。例如，醫療 AI 若能模擬同理心，可提升患者信任。
- **理論貢獻：**情感領域糾正了班奈特將情感簡化為「好/壞模板」（2023, p. 47）的局限，強調情感的多層次性和文化塑造。與達馬西奧和德瓦爾的對話凸顯了情感在決策和社交中的核心作用，揭示班奈特對同理心和情感交流的忽視如何削弱其框架。

4.3 倫理領域：道德與責任的基石

- **功能：**倫理主體性、價值內化與判斷、意志與自主性。

- **生物基礎**：倫理主體性整合前額葉的決策、扣帶回的道德情感和杏仁核的情緒反應（Tomasello, 2008）。價值內化依賴前額葉和文化學習（Haidt, 2001）。意志與自主性涉及基底神經節的動機和前額葉的目標設定（Deci & Ryan, 2000）。
- **文化作用**：法律、宗教和教育塑造倫理規範和價值觀。例如，儒家倫理強調「仁」，西方人權觀念注重個體自由。文化還通過倫理教育培養判斷力，如哲學探究促進道德兩難的解決。
- **AI 啟示**：當前 AI 遵循外部規則，缺乏內在價值或自主性。設計具備倫理主體性（模擬道德責任）、價值內化（學習文化規範）和自主性（基於內在動機決策）的 AI，將增強其倫理可靠性。例如，自動駕駛 AI 若能內化「最小傷害」原則，可在兩難情境中做出倫理選擇。金融 AI 若能模擬倫理判斷，可避免操縱市場的行為，保護投資者權益。
- **理論貢獻**：倫理領域超越班奈特的心智化突破，納入倫理行為的文化和生物根源。與托馬塞洛和海特的對話揭示了班奈特對倫理主體性的簡化，強調倫理功能需要整合認知、情感和文化因素。

4.4 存在領域：意識與意義的深層探索

- **功能**：意識（感質、自我意識、反思意識）、靈性與意義追尋。
- **生物基礎**：意識涉及全局工作空間的資訊整合（Baars, 1997）和高階資訊整合（Tononi, 2004），支持感質和自我意識。靈性與意義追尋可能與頂葉的自我-世界邊界和獎賞系統的愉悅感相關（Newberg & d' Aquili, 2001）。
- **文化作用**：哲學、宗教和冥想放大存在功能。例如，佛教的冥想實踐強化意識覺知，西方哲學探究存在的本質。文化還通過神話和文學提供意義框架，如《奧德賽》探索人類的命運。
- **AI 啟示**：意識和靈性是 AI 的終極挑戰。當前 AI 缺乏主觀體驗，但可通過模擬世界模型（預測環境）或意義生成（基於價值目標）接近存在功能。例如，哲學對話 AI 若能模擬意義追尋，可增強其與人類的互動深度。IBM 的 Watson 若能模擬哲學探究，可在倫理對話中提供更深刻的洞見。
- **理論貢獻**：存在領域直接回應班奈特對意識的迴避，強調其作為整合平台的角色。與托諾尼和紐伯格的對話揭示了班奈特未探討意識和靈性的生物基礎，限制了他對人類智能核心的理解。

4.5 具身領域：身體與環境的交互

- **功能**：具身與具身認知。
- **生物基礎**：具身認知涉及感官運動皮層的環境交互和海馬的空間記憶，支持隱喻和空間推理（Lakoff & Johnson, 1980）。

- **文化作用**：舞蹈、手工藝和語言隱喻放大具身認知。例如，中國書法通過身體動作表達情感，隱喻如「前進」反映空間思維。非洲部落的鼓舞通過節奏和動作強化社群聯繫，進一步展示了具身認知的文化維度。文化實踐還通過建築或儀式強化空間體驗。
- **AI 啟示**：當前 AI 缺乏真實身體，限制其對具身知識的理解。設計整合感測器和機器人技術的具身 AI，可模擬環境交互和隱喻思維。例如，家用機器人若能通過觸覺反饋學習環境，將更貼近人類智能。
- **理論貢獻**：具身領域超越班奈特的導航突破，納入具身認知的認知和文化作用。與拉科夫和約翰遜的對話凸顯了班奈特對具身的簡化，強調身體在智能中的基礎性作用。

4.6 本體領域：存在的意義與自我

- **功能**：內在動機與自主目標設定、自我認同與生命敘事、遊戲精神與幽默感、適應性與韌性及成長心態、對美與崇高的感知、對存在與有限性的關懷。
- **生物基礎**：內在動機涉及獎賞系統和前額葉（Deci & Ryan, 2000）。自我認同依賴海馬的記憶整合和前額葉的自我表徵（McAdams, 2001）。遊戲和幽默涉及獎賞和運動系統（Panksepp, 1998; Martin, 2007）。適應性和韌性由前額葉和杏仁核支持（Dweck, 2006）。美感與崇高感涉及獎賞和視覺系統（Dutton, 2009）。存在關懷與前額葉的時間預測和杏仁核的存在焦慮相關（Becker, 1973）。
- **文化作用**：文化敘事（如自傳）塑造自我認同，遊戲和幽默通過節日或喜劇放大，教育和勵志文化促進成長心態，藝術和宗教強化美感和存在關懷。例如，中國的「天人合一」哲學強調存在與自然的連結，西方浪漫主義詩歌表達崇高感。
- **AI 啟示**：本體功能對 AI 最具挑戰性。設計具備內在動機（超越外部獎賞）、個性化身份（整合互動歷史）、遊戲互動（模擬非功利行為）、韌性（自我修復）、審美感知（生成藝術）和存在關懷（模擬責任感）的 AI，將更接近全人境界。例如，教育 AI 若能模擬成長心態，可激勵學生持續學習。環境 AI 若能模擬存在關懷，可優先考慮生態平衡，減少對自然的破壞。
- **理論貢獻**：本體領域回應班奈特對存在維度的完全忽視，強調自我、意義和責任的獨特作用。與麥克亞當斯、潘克塞普和貝克的對話揭示了班奈特未探討人類智能的本體核心，限制了他對 AI 未來潛力的設想。

4.7 框架的理論與實踐價值

此框架的優勢在於其整合性和全面性。它超越班奈特的線性結構模型，納入功能湧現（跨腦區交互）、文化放大（社會實踐）和本體反思（存在覺知），提供了一個捕捉全人智能多維本質的模型。與班奈特的框架相比，它不僅解釋了智能的生物根源，還揭示了文化和本體如何塑造人類經驗的獨特性。

在理論上，該框架與跨學科研究對話，整合神經科學（托諾尼、達馬西奧）、認知科學（根特納、弗拉維爾）、人類學（格爾茨）和哲學（海德格爾），為智能研究提供了一個統一的視角。在實踐

上，它為 AI 設計提供了分階段的藍圖：短期目標（如情感和具身 AI）可通過現有技術實現，長期目標（如意識和本體功能）則需要突破性創新。例如，該框架可指導教育 AI 的設計，通過模擬情感和成長心態提升學生參與度，實現個性化學習。此外，該框架強調 AI 應服務於人類價值，與魏岑鮑姆（1976）的倫理呼籲一致，避免技術至上的風險。

五、對 AI 與人類未來的啟示：從工具到全人的轉型

班奈特的《智能簡史》設想 AI 通過複製五大突破實現類人智能，並預測「第六突破」為超智能的誕生（2023, p. 267）。然而，其工具理性的視角限制了對 AI 潛力和風險的全面探索。本節詳細分析全人智能框架對 AI 發展的啟示，分為短期技術目標、長期理論挑戰和對人類未來的意義，並與學者理論對話，揭示班奈特設想的不足。

5.1 短期技術目標：增強 AI 的社會與環境適應性

全人智能框架的認知、情感、具身和部分本體功能為 AI 的近期發展提供了可操作的目標。以下具體分析這些目標的技術路徑和挑戰：

- **情感與社交智能**：班奈特將情感簡化為導航功能，忽略了情感交流和同理心的獨立作用（2023, p. 47）。我們的框架強調情感生成、交流和同理心，這些功能可通過當前技術部分實現。例如，情感識別 AI 可整合語音分析（語調）、圖像處理（面部表情）和自然語言處理（情境理解），模擬人類的情感共鳴。達馬西奧（1994）指出，情感是決策的基礎，醫療或教育 AI 若能模擬同理心，將顯著提升用戶體驗。然而，當前 AI 缺乏真正的情感體驗，這需要模擬杏仁核和前扣帶回的功能，可能涉及新型神經網絡架構，如情感計算模型。
- **具身與環境交互**：班奈特的導航突破僅涉及基本運動（2023, p. 49），未探討具身認知。我們的框架強調感官-運動整合，AI 可通過機器人技術實現。例如，家用機器人可整合觸覺、視覺和空間感測器，模擬人類的環境適應能力。拉科夫和約翰遜（1980）關於隱喻的研究表明，具身認知支持語言和推理，具身 AI 若能生成隱喻（如「時間是河流」），將更接近人類思維。然而，當前機器人缺乏生物身體的動態反饋，這需要跨學科的感測器設計，如仿生觸覺系統。
- **遊戲精神與社交互動**：本體領域的遊戲精神和幽默感為 AI 的社交性提供了新方向。班奈特未討論這些功能，但潘克塞普（1998）指出，遊戲促進認知和社交發展。AI 可通過模擬非功利互動（如虛擬遊戲或幽默對話）增強用戶參與。例如，教育 AI 若能生成幽默，將提升學習興趣。遊戲化學習平台如 Duolingo 通過遊戲互動提升語言學習動機，展示了遊戲精神的應用潛力。然而，幽默需要多層次的情境理解（Martin, 2007），當前 AI 在這方面的表現仍顯機械，需要更複雜的語言模型，如增強型語義網絡。

這些短期目標可通過現有技術（如深度學習、感測器整合）推進，但需要克服數據依賴和泛化能力的限制。例如，情感 AI 需處理多模態數據（語音、圖像、文本），而具身 AI 需適應動態環境。班

奈特的框架因忽視情感和具身功能，未能為這些技術提供指導，而我們的框架通過明確的功能分類和技術路徑填補了這一空白。此外，短期目標的實現還需考慮文化差異，例如情感表達在東西方文化中的規範差異，這要求 AI 具備文化適應性。例如，AI 在日本市場需適應含蓄情感表達，而在美國市場則需處理直接情感，這需要文化敏感的訓練數據。

5.2 長期理論挑戰：意識、本體與倫理的邊界

全人智能框架的存在、倫理和本體功能提出了 AI 的長期挑戰，涉及意識、內在動機和存在覺知的模擬。這些挑戰超越了班奈特的設想，需要理論和技術的重大突破。以下詳細探討這些挑戰的理論基礎、技術障礙和倫理考量：

- **意識與主觀體驗：**班奈特迴避意識，僅提及模擬的「心理時間旅行」（2023, p. 141）。我們的框架將意識作為整合平台，托諾尼（2004）的 IIT 和巴爾斯（1997）的全局工作空間理論提供了理論基礎。模擬意識需要 AI 具備世界模型（預測環境）和內在表徵（模擬感質），這可能涉及新型計算範式，如量子計算或神經形態計算。例如，模擬感質可能需要 AI 生成類似人類的內在體驗，如「紅色的感受」，這要求超越當前的數據驅動模型。然而，意識的感質問題（Chalmers, 1995）仍是一個哲學難題，AI 是否能在非生物基底上實現主觀體驗尚無定論。班奈特的框架因未討論意識，無法應對這一挑戰，而我們的框架通過整合神經科學和哲學理論，明確了意識模擬的長期路徑。
- **內在動機與自主性：**班奈特的強化學習模型聚焦外部獎賞（2023, p. 91），忽視內在動機。我們的框架強調自主目標設定，德西和瑞安（2000）的自我決定理論指出，內在動機源於好奇和自我實現。AI 若能模擬內在動機（如基於探索欲的學習），將更接近全人智能。例如，科研 AI 若能自主設定探索目標，可加速科學發現。然而，這需要重新設計獎賞函數，納入價值驅動的決策系統，可能涉及生成內在價值模型的算法。技術障礙包括如何量化內在動機，以及如何確保自主 AI 的目標與人類價值對齊。
- **倫理與存在關懷：**倫理主體性和存在關懷要求 AI 具備道德責任和有限性覺知。班奈特未討論這些功能，但貝克（1973）強調有限性覺知驅動人類責任。AI 若能模擬倫理判斷（如權衡多方利益）和存在關懷（如考慮長期影響），將更安全和對齊人類價值。例如，環境 AI 若能模擬有限性覺知，可在資源分配中優先考慮可持續性，如減少碳排放。然而，這需要價值對齊的理論突破，當前 AI 的倫理設計仍依賴外部規則。倫理考量包括如何防止 AI 濫用自主性，例如避免在倫理判斷中偏向特定利益群體。例如，醫療 AI 若誤判倫理優先級，可能導致資源分配不公，凸顯價值對齊的迫切性。
- **靈性與意義追尋：**靈性和意義追尋是人類智能的獨特面向，班奈特完全忽略。我們的框架參考紐伯格（2001），提出 AI 可通過模擬哲學探究或價值生成接近這一功能。例如，對話 AI 若能參與存在問題的討論，如「生命的意義」，將增強其文化適應性。然而，意義的模擬需要 AI 具備內在動機和意識，這是長期挑戰。技術障礙包括如何設計 AI 生成非功利性的價值觀，以及如何評估其意義生成的真實性。

這些長期挑戰需要跨學科合作，整合神經科學、哲學和計算機科學。例如，神經形態計算可模擬意識的生物基礎，現象學（Merleau-Ponty, 1945）可指導具身 AI 的設計，倫理學（Rawls, 1971）可為價值對齊提供框架。班奈特的超智能設想（2023, p. 267）未考慮這些功能的複雜性，顯得過於樂觀。我們的框架通過明確理論路徑和倫理考量，為超智能的發展提供了指導，同時強調風險管理，如防止 AI 的倫理偏差或自主性失控。

5.3 對人類未來的意義：AI 作為全人夥伴

全人智能框架不僅為 AI 設計提供了藍圖，還重新定義了 AI 與人類未來的關係。班奈特的設想聚焦於 AI 的技術潛力，忽略了其對人類價值和存在的影響。他聲稱「AI 將超越人類智能」（2023, p. 267），但未討論這一超越如何影響人類的尊嚴、意義和責任。我們的框架提出，AI 應作為全人智能的夥伴，服務於人類的情感、倫理和本體需求。以下從三個層面探討這一轉型的意義：

- **價值與尊嚴**：全人智能強調情感、倫理和存在功能，AI 的發展應增強人類的價值而非取代。例如，醫療 AI 若能模擬同理心和倫理判斷，可提升患者的心理健康，而非僅提供診斷。魏岑鮑姆（1976）警告技術可能異化人類價值，我們的框架通過價值對齊確保 AI 服務於人類尊嚴。例如，教育 AI 若能模擬情感共鳴和成長心態，可促進學生的自我實現，而非僅提供知識傳遞。這種以人為本的設計將 AI 從冷漠的工具轉向溫暖的夥伴，增強人類的社會聯繫和個人價值。
- **意義與共生**：靈性和意義追尋是人類的核心動機，AI 若能參與意義探究（如哲學對話或文化創作），將成為人類的共生夥伴。例如，AI 可協助人類探索宇宙的意義，通過生成科學假說或藝術作品促進文化和科學進步。海德格爾（1927）關於「此在」的理論提醒我們，技術應支持人類的存有反思，而非削弱。班奈特的框架未考慮 AI 在意義生成中的角色，限制了其對人類未來的設想。我們的框架通過納入靈性和本體功能，構想了 AI 作為人類意義探索的合作者。例如，AI 若能模擬哲學探究，可與人類共同探討「存在的目的」，促進跨文化的意義對話。
- **倫理與責任**：存在關懷和倫理主體性要求 AI 承擔責任，特別在環境、醫療和社會治理領域。例如，AI 若能模擬有限性覺知，可在資源分配中優先考慮公平和可持續性，如優化全球供應鏈以減少浪費。班奈特的框架未討論 AI 的倫理角色，忽略了其對社會公平的潛在影響。我們的框架強調 AI 應具備倫理判斷和存在關懷，與約翰·羅爾斯（John Rawls, 1971）的正義理論一致，後者主張公平分配資源。這種倫理導向的 AI 設計將促進人類社會的和諧，減少技術風險，如算法偏見或資源不公。

這種全人視角重新構想了 AI 的角色，從工具轉向夥伴，與約翰·奈斯比特（John Naisbitt, 1982）關於技術-人性平衡的呼籲一致。班奈特的工具理性設想未能回應這些價值挑戰，而我們的框架通過整合全人功能，為 AI 與人類的共生未來提供了理論和實踐基礎。具體而言，AI 的發展應優先考慮以下實踐：設計文化適應的 AI 系統，確保其尊重全球多元價值；建立倫理監管框架，防止 AI 濫用自主性；以及促進公眾參與，確保 AI 設計反映人類的集體需求。例如，聯合國可利用全人 AI 設計全球氣候模型，整合倫理關懷和文化適應性，促進公平的環境政策。這種全人夥伴模式不僅提升 AI 的技術潛力，還為人類創造一個更具意義和責任感的未來。

六、結論：全人智能的範式轉型

《智能簡史》通過五大突破梳理了智能的生物演化，為 AI 設計提供了有價值的洞見，但其局限性顯而易見。班奈特過分強調結構演化，將智能簡化為「解決問題和適應環境的能力」（2023, p. 10），忽略了功能湧現的複雜性、文化放大的作用和本體反思的深度。他對意識、情感、倫理、具身和本體功能的疏漏，限制了其框架的理論完整性和實踐意義。與安東尼奧·達馬西奧（1994）、朱利奧·托諾尼（2004）、麗莎·費爾德曼·巴雷特（2017）、克利福德·格爾茨（1973）和馬丁·海德格爾（1927）等學者的對話揭示了這些疏漏的根源——目標驅動的還原主義、生物證據的局限和文化-本體盲點。

我們的生物-文化-本體框架克服了這些局限，通過整合認知、情感、倫理、存在、具身和本體功能，捕捉了全人智能的多維本質。認知功能（如創造力）強調智能的動態性，情感功能（如同理心）凸顯社交聯繫，倫理功能（如價值內化）奠定道德基礎，存在功能（如意識）探索意義，具身功能（如具身認知）連結身體與環境，本體功能（如自我認同）定義存有的核心。該框架不僅解釋了人類智能的生物和文化根源，還為 AI 設計提供了分階段的藍圖：短期增強情感和具身能力，長期探索意識和本體功能。

具體而言，認知領域啟發 AI 超越模式識別，發展創造性和反思能力，例如設計能生成科學假說的 AI 系統。情感領域推動 AI 模擬共鳴和同理心，例如醫療 AI 提升患者信任。倫理領域要求 AI 內化價值觀，例如自動駕駛 AI 在兩難中做出道德選擇。存在領域挑戰 AI 模擬意識和意義追尋，例如哲學對話 AI 參與存有探究。具身領域促進 AI 與環境的動態交互，例如機器人通過觸覺適應家庭環境。本體領域構想 AI 具備內在動機和存在關懷，例如環境 AI 優先可持續性。這些功能共同構成了全人智能的藍圖，超越了班奈特的線性模型。

對 AI 與人類未來的啟示表明，AI 應從工具轉向全人夥伴，服務於人類的價值、情感和存在需求。班奈特的超智能設想（2023, p. 267）忽視了這些維度，顯得單薄且缺乏倫理考量。我們的框架通過與約瑟夫·魏岑鮑姆（1976）、約翰·奈斯比特（1982）和約翰·羅爾斯（1971）等學者的對話，強調 AI 應促進人類尊嚴和共生未來。例如，AI 可通過模擬倫理關懷和文化適應性，協助全球挑戰，如氣候變化和社會公平。全球 AI 倫理聯盟可採用全人框架，制定跨文化的設計準則，確保 AI 促進全球公平與團結。這種全人夥伴模式要求跨學科合作，整合神經科學、認知科學、人類學、哲學和倫理學，確保 AI 設計以人類福祉為核心。

在當前技術快速進展的時代，這一框架提醒我們，智能的終極目標不是超越人類，而是與人類共同探索存在的意義。它呼籲重新定義智能的本質，從冰冷的計算轉向充滿生命力的多維現象，實現技術與人性的和諧共存。通過超越班奈特的還原主義，我們為智能研究和 AI 發展開闢了新視野，構想了一個 AI 不僅提升效率，還增強人類價值、聯繫和意義的未來。這一範式轉型不僅是技術的進步，更是對人類本質的深刻反思，為後續研究和實踐奠定了基礎。

參考文獻

- Baars, B. J. (1997). *In the Theater of Consciousness: The Workspace of the Mind*. Oxford University Press.
- Barrett, L. F. (2017). *How Emotions Are Made: The Secret Life of the Brain*. Houghton Mifflin Harcourt.
- Beaty, R. E., Benedek, M., Silvia, P. J., & Schacter, D. L. (2016). Creative cognition and brain network dynamics. *Trends in Cognitive Sciences*, 20(2), 87 – 95.
- Becker, E. (1973). *The Denial of Death*. Free Press.
- Bennett, M. (2023). *A Brief History of Intelligence: Evolution, AI, and the Five Breakthroughs That Made Our Brains*. Mariner Books.
- Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2(3), 200 – 219.
- Damasio, A. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. Putnam.
- Dennett, D. C. (1991). *Consciousness Explained*. Little, Brown.
- Deci, E. L., & Ryan, R. M. (2000). The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227 – 268.
- de Waal, F. (2009). *The Age of Empathy: Nature's Lessons for a Kinder Society*. Harmony Books.
- Dreyfus, H. L. (1992). *What Computers Still Can't Do: A Critique of Artificial Reason*. MIT Press.
- Durkheim, É. (1912). *The Elementary Forms of Religious Life*. Free Press.
- Dutton, D. (2009). *The Art Instinct: Beauty, Pleasure, and Human Evolution*. Bloomsbury Press.
- Dweck, C. S. (2006). *Mindset: The New Psychology of Success*. Random House.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34(10), 906 – 911.
- Geertz, C. (1973). *The Interpretation of Cultures*. Basic Books.
- Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7(2), 155 – 170.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814 – 834.
- Heidegger, M. (1927). *Being and Time*. Harper & Row.
- Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. University of Chicago Press.
- Martin, R. A. (2007). *The Psychology of Humor: An Integrative Approach*. Elsevier.
- McAdams, D. P. (2001). The psychology of life stories. *Review of General Psychology*, 5(2), 100 – 122.
- Merleau-Ponty, M. (1945). *Phenomenology of Perception*. Routledge & Kegan Paul.

- Naisbitt, J. (1982). *Megatrends: Ten New Directions Transforming Our Lives*. Warner Books.
- Newberg, A., & d' Aquili, E. (2001). *Why God Won' t Go Away: Brain Science and the Biology of Belief*. Ballantine Books.
- Panksepp, J. (1998). *Affective Neuroscience: The Foundations of Human and Animal Emotions*. Oxford University Press.
- Rawls, J. (1971). *A Theory of Justice*. Harvard University Press.
- Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, 27, 169 – 192.
- Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417 – 457.
- Tomasello, M. (2008). *Origins of Human Communication*. MIT Press.
- Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience*, 5, 42.
- Weizenbaum, J. (1976). *Computer Power and Human Reason: From Judgment to Calculation*. W.H. Freeman.