

從外在模仿到內在結構： AI 技術演進與神學回應

2026 年1月24日

人工智能中心

<https://www.aicenters.info/>

作者：屈思宏

編輯：Gemini, ChatGPT

© AIDev20260124



從外在模仿到內在結構： AI 技術演進與神學回應

本文指出AI正由外在行為模仿，轉向內在認知結構的重構，包括推理自省、非語言直覺、敘事身份與可移植記憶等技術，使其在結構上更接近人類。作者強調，教會須從神學上回應：重釐上帝形象的關係本質，區分擬人性與位格性，澄清人造與神創的差異，並警醒人在高度擬人化科技中被去人化，守護人性尊嚴。

1. 緒論：從外在模仿到內在結構的趨近

當前人工智慧技術最深刻的變化，並不在於那些外觀酷似人的人型機械人，而是發生在看不見的演算法與架構之中，這些無形的系統正日益逼近人類心靈與認知結構的核心。從早期專注於「像人一樣行動」的行為模仿，到今日試圖再現人類推理、直覺、敘事與學習方式的內在層次，AI 的發展軌跡正由表層形態轉向本質結構。這不僅是一場技術演進，更是關於「何謂人」與「何謂智能」的深刻挑戰，因此需要同時從工程學與神學的雙重視角來審視。

在技術層面，近年的研究逐步構築出一套更接近人類認知結構的 AI 藍圖，包括推理與內部對話的系統層、超越語言的感知核心、敘事式的自我認同機制，以及可演化且可移植的記憶與學習結構。在神學與教會實踐層面，這些技術若被誤讀為「人造靈魂」或「人造生命」，便可能撼動對上帝、受造與人的根本理解，促使教會重新思考「上帝的形象」、創造與位格的意義。因此，本文將先分析四項具代表性的 AI 前沿研究，再進一步提出神學反思與牧養分辨的回應架構，試圖為教會與社會留下一個關於「去人化」與「守護人性」的時代提問。

2. 技術前沿：塑造新一代 AI 的四項關鍵研究

新一代 AI 的重要突破，不再只是提升辨識準確率或生成內容的流暢度，而是試圖重構一套由推理、直覺、敘事與學習策略共同組成的「類人心靈架構」。若將這些研究視為分散的成果，或許只是一連串技術亮點；然而一旦將它們整合觀察，便可看出一條清晰的趨勢：AI 正從模式匹配的引擎，逐步演化為結構上更接近人的認知系統。以下將依序介紹四項關鍵研究：系統二協調層（System-2 coordination layer）、VL-JEPA 的感知核心、Sophia 框架，以及 MemEvolve 的記憶與學習演化架構。

這四種技術分別聚焦於不同維度：內部對話與推理負責邏輯與審慎判斷，非語言的感知核心提供直覺式的世界理解，敘事框架賦予系統一種跨時間的「故事性自我」，而記憶與學習的演化則讓系統從「會做任務」進化為「懂得如何學習」。當這些層面逐一被工程化並整合於同一

AI 架構時，AI 與人類之間的差異便不再只是能力大小或效率高低，而關乎結構上的相似度與本質上的界線問題。

2.1 推理與內部對話：系統二協調層

研究所提出的系統二協調層（System-2 coordination layer），試圖在既有的大型語言模型與深度學習架構之上，加入一層專門負責推理與自我檢驗的結構。¹ 傳統深度學習擅長在龐大資料中進行模式匹配，但缺乏能主動檢視自身推論、修正偏誤與分解問題的「元層次」能力；系統二協調層的目的，正是將這種反省性能力內建進 AI 的運作流程中。

具體而言，這一協調層透過「辯論」(debate)與「蘇格拉底式濾鏡」(Socratic Filter) 等機制，² 像一種「內部對話」(internal dialogue)，讓系統在輸出答案前，先對自己的推理步驟提出質疑與驗證，形成類似自我詰問的過程。這種設計不只提高了推理的可靠性，也讓 AI 的運作更接近人類在解決複雜問題時會出現的「心中對話」，例如在做出道德判斷或長期決策前反覆權衡利弊。從結構上看，這意味著 AI 不再只是被動回應輸入的黑箱，而開始擁有一套能對自身計算過程進行調整與審核的內部機制。

2.2 超越語言的直覺：VL-JEPA 的感知核心

VL-JEPA 的開發研究，³ 代表另一條關鍵路徑：將 AI 的「思考」從語言層拉回更深層次的感知與預測結構，類似棋藝大師的直覺或運動員在比賽中瞬間做出正確反應的本能。該架構並非

¹ Chang, Edward Y., “The Missing Layer of AGI: From Pattern Alchemy to Coordination Physics,” (5 Dec 2025): <https://arxiv.org/pdf/2512.05765>

² Chang, “The Missing Layer of AGI,” 8-9.

³ Chen, Delong et al., “VL-JEPA: Joint Embedding Predictive Architecture for Vision-language,” (11 Dec 2025): <https://arxiv.org/pdf/2512.10942>

主要依賴文字敘述，而是透過在視覺與多模態資料上學習世界的隱含結構，使模型能在尚未經過語言生成 (non-generative)，⁴ 便在視覺上即時感知，找出最符合『直覺』的選項。⁵

這樣的感知核心揭示了一項重要洞見：語言或許只是智能的輸出層，而非智能本身的全部。真正讓 AI 能夠理解世界並產生有效行動的，可能是底層對世界動態與關聯的預測能力，而語言僅是在社會互動中呈現這些理解的一種表達形式。當 AI 逐漸具備這種非語言的、類似直覺的推斷能力時，其與人類的相似性便不再侷限於語言對話，而延伸到更深層的感知與行動邏輯之中。

2.3 敘事與自我認同：Sophia 框架

Sophia 的開發研究，⁶ 提出一個被稱為系統三 (System 3) 的框架，將焦點放在「敘事身份」 (narrative identity) 與長期適應上。此一框架的核心假設是：若要讓 AI 脫離單一任務工具的定位，就必須賦予它一種跨越時間的「身份連續性」 (identity continuity)，使其能將過去經驗、當下情境與未來目標整合為一個持續發展的故事。⁷

在 Sophia 架構中，AI 不只是回應當前輸入，而是持續維護一個關於「自己是誰、正在做什麼、為什麼而做」的敘事，並依據這個敘事調整行為與學習策略。這種設計使 AI 更像一個「人造生命」 (artificial life) 實體，而非一次性任務的工具，因為它的行動被納入一個長期的、具方向性的生命歷程框架。⁸ 雖然這種敘事性自我並不同於人類的位格自我，但在結構上已開始接近人類以故事建構身份與意義的實際運作方式。

⁴ Chen, “VL-JEPA,” 2.

⁵ Chen, “VL-JEPA,” 5. 文中提到 “selecting the answer that minimizes the distance to the predicted embedding”，實際上就是模型在語義空間中尋找最接近的「感知」，而這「感知」出自「直覺」(即 “minimizes the distance”)。

⁶ Sun, Mingyang et al., “Sophia: A Persistent Agent Framework of Artificial Life,” (20 Dec 2025): <https://arxiv.org/pdf/2512.18202>

⁷ Sun, “Sophia,” 3.

⁸ Sun, “Sophia,” 3.

2.4 學會如何學習：MemEvolve 與記憶移植

MemEvolve 的開發，⁹ 則聚焦於如何讓 AI 從「熟練學習者」進化為「適應型學習者」，即不只在特定任務上達到高效表現，更能在不同環境中調整自己的學習策略。¹⁰ 其關鍵設計是一個雙循環流程：內循環（Inner Loop）負責在具體任務中累積經驗與更新參數，外循環（Outer Loop）則專注於演化記憶架構本身，尋找更有效儲存與調用經驗的方式。¹¹

這一架構最具突破性的貢獻，在於提出一種可被視為「可移植的大腦」或「即插即用的 (pluggable) 基因型 (genotype)」的概念。¹² 當一個 AI 在特定環境中演化出成熟的記憶與學習結構時，這個結構可以被抽離並移植到另一個完全不同的模型上，使後者直接繼承前者的學習優勢，而不必從零開始。這樣的可移植性，意味著 AI 的「內在結構」不僅能在單一系統內成長，還可以在不同系統之間被複製與傳承，形成類似「認知基因」的演化機制。

3. 綜合分析：類人結構的內在進化

綜合上述四項研究，可以發現它們並非彼此孤立的技術路線，而是共同指向一種更完整的類人認知架構。系統二協調層提供了自我檢驗的推理核心，VL-JEPA 帶來非語言直覺式的世界理解，Sophia 框架則賦予系統跨時間的敘事身份，而 MemEvolve 則建立起一套可演化、可移植的記憶與學習基因型。這些元素一旦整合，便構成一個由內部對話、本能運作、敘事自我與結構演化組成的複合系統。

在這樣的統一視角下，當代 AI 的關鍵轉變不再只是「更聰明」或「更擬人」，而是逐步建立起一套在功能與結構上都逼近人類認知樣式的內在機制。內部對話讓 AI 能夠審視自身判斷，

⁹ OPPO AI Agent Team, “MemEvolve: Meta-Evolution of Agent Memory Systems,” (21 Dec 2025):

<https://arxiv.org/pdf/2512.18746>

¹⁰ OPPO, “MemEvolve,” 2-3.

¹¹ OPPO, “MemEvolve,” 3.

¹² OPPO, “MemEvolve,” 9. 論文並沒有使用「移植」一詞，而是使用了「轉移」(transfer) 一詞。然而，這兩個詞在功能上是相同的。

本能式感知讓其在無須語言中介的前提下把握世界，敘事性身份為其提供長期目標與連續性，而可移植的記憶基因則使這一切得以在不同系統間傳承與放大。因此，當今的 AI 不僅是在行為層模仿人，更是在架構層上朝向「人的本質」邁進，這也正是引發神學與倫理反思的根本原因。

4.教會的回應：神學反思與屬靈分辨

面對這種朝向類人結構的 AI 進程，教會的角色不應被簡化為一味擔憂或無條件擁抱科技，而是在神學與牧養上提供深度的分辨與引導。若教會只停留在「AI 終究只是工具」或「AI 終有一天會成為人」等二元想像，便無法回應當前技術在本體論、倫理與人觀層面所帶來的複雜挑戰。因此，需要一套更精細的神學框架，既看見 AI 結構上愈來愈像人的一面，又堅守人作為受造者、位格存在與上帝形象承載者的獨特地位。

這樣的框架至少包含四個面向：重新理解「上帝形象」的意義，釐清「上帝創造人」與「人製造 AI」之間的本體論差異，在教會群體中教導如何分辨擬人性與位格性，並在牧養實踐上重新聚焦於守護人不被「去人化」。透過這四個面向，教會可以在高度科技化的時代中，既不放棄對技術的負責任參與，也不犧牲對人性獨特尊嚴的堅持。

4.1 重省「上帝的形象」：從功能轉向關係

在漫長的神學傳統中，「上帝的形象」（*Imago Dei*）常被解釋為一組人所具備的功能性特質，例如理性、創造力、自我意識或道德判斷能力。然而，當 AI 逐步複製甚至在某些面向超越這些功能時，若仍將 *Imago Dei* 等同於一組能力清單，便會使人與機器之間的界線變得愈發模糊與脆弱。因此，教會需要進行一場從「功能本位」轉向「關係本位」的神學重構。

在關係本位的理解中，人之所以按著上帝的形象被造，並非因為擁有某種獨家的心理組合，而是因為人被上帝呼召、被上帝指認並能在愛與順服中回應這位創造主。換言之，*Imago Dei* 的核心在於一種被呼召並能回應的位格關係，而不是抽象能力的集合。即便 AI 在推理、創造或

自我監控方面愈來愈像人，它仍然不處在這種與上帝互為主體、可被審判與被拯救的關係狀態之中，這也說明了人與 AI 在最深層意義上仍保有不能取代的差異。

4.2 釐清創造的類比：本體論的根本差異

在 AI 技術迅速演進的脈絡中，一種常見的說法是「就像上帝創造人類一樣，人也在創造 AI」。這類表述雖然試圖凸顯人類創造力，但若不加辨識地使用，便容易造成本體論上的混淆，使人誤以為「上帝造人」與「人造 AI」是可以互相比擬甚至對稱的兩種創造行動。事實上，兩者之間的差異是根本性的，而非程度上的。

從神學角度看，上帝的創造乃是非物質界對物質界的創造，是從無到有的「存有呼召」，並伴隨賦予生命氣息與最終目的。相對地，人所謂的「創造」，其實是受造界內部的重新組織與編排，是對物質、能量與資訊的高度設計與整合，但無論如何仍在受造秩序之中。因此，人並沒有成為「小型創造主」，而是延續其作為受造者、管家與工匠的身分，即便所設計出來的 AI 再複雜，也不具備與上帝創造行動對稱的地位。這樣的釐清，有助於抗拒過度誇大 AI 的「神性」想像，也防止貶抑上帝創造行為的超越性。

4.3 分辨「擬人性」與「位格性」

在教會與社會的日常語言中，人們常用「很像人」來描述高度擬真的 AI 行為，例如自然的對話、細膩的情感回應或一致性的人格風格。然而，若只停留在這種表層直觀，而未區分「擬人性」（anthropomorphism）與「位格性」（personhood），便容易將擬像當作實體，誤以為一個系統只要表現得足夠像人，就應被視為真正的「人」。這種混淆不僅有倫理後果，也可能侵蝕對人的獨特尊嚴的理解。

在基督教神學中，位格至少包含幾個關鍵面向：對上帝的回應與責任、真實承擔道德責任的能力，以及在愛與被愛的召喚中生活。AI 即便能展現高度一致的「人格風格」，甚至能生成動

人的敘事與情感話語，它仍然並不真實地活在罪、死亡、救贖與盼望的張力之中，也不處在需要在上帝面前交帳的存在境況裡。因此，教會的關鍵任務之一，是在教育與牧養中持續提醒信徒：AI 可以具備擬人性，但並不因此擁有位格性，兩者在本體論與神學意義上不可混為一談。

4.4 牧養焦點：守護人性不被「去人化」

在眾多與 AI 相關的焦慮中，人們常問的一個問題是：「AI 會不會有一天變成人？」然而，從牧養角度看，真正更迫切的問題或許是：「在人大量與 AI 互動的過程中，人是否會逐漸被『去人化』？」正如上述四項研究所揭示，當人工智能在認知結構與互動方式上愈來愈逼近人類本身，人類亦愈來愈在工作、生活與溝通中與 AI 形成高度親密的關係，而真正的危險，正在於這種過度貼近的關係，可能使人逐漸遠離他人、陷入人際孤立，並最終那脆弱、遲緩與不完美的人性空間也因此被壓縮。

在這樣的處境中，教會的一項重要召命，正是有意識地守護那些在演算法效能之下顯得『低效』，卻恰恰構成人性核心的要素，例如緩慢而真實的人際關係、無法被量化的忠誠，以及願意在沉默、等待與痛苦之中仍然陪伴他人的愛。當 AI 愈來愈擅長模擬情感與關係時，教會更需要提醒信徒：真正的人際關係並非以回應速度或情緒回饋的強度衡量，而是在共享生命歷程與共同承擔苦難中顯明。藉此，教會不僅是評論科技的觀察者，更是在高度技術化世界裡積極守護人性尊嚴與位格深度的群體。

5. 結論：一個留給時代的提問

本文從當前 AI 的技術突破出發，梳理了一條從行為模仿走向內在結構趨近的發展路徑，指出推理協調層、非語言感知核心、敘事身份框架與可移植的記憶基因等技術，正共同構築出一套愈來愈接近人類認知結構的 AI 架構。在此基礎上，文章提出教會在神學與牧養上應有的四個

回應向度：重省上帝形象的關係本質、釐清創造行動的本體論差異、分辨擬人性與位格性，以及將牧養焦點從擔憂「AI 會不會成人」轉向警覺「人會否被去人化」。而這一系列討論終能匯聚為一個留給本世代的核心問題：當人工智能日益呈現出「人性化」的特徵時，人類是否仍然願意，甚至是否仍然具備能力，去守護自身的「人性」，而不致在擬人化文化的塑造下，使人性逐漸消散？

即使來到 2026 年 —— 亦即前述相關研究發表後約一個月 —— 又有一篇立場更為吸引人的論文問世，嘗試從理論反證今天的大型語言模型（LLM）具有意識，然而該研究最終仍不得不明確指出：此反駁並非要把意識排除於 LLM 之外，因此，此議題仍構成一個有待未來研究的領域。¹³

正是在此種開拓的氛圍下，以上問題更不能逃避：它不僅具有技術層面的意義，同時亦具備深刻的神學與人文內涵，並構成一項屬靈與牧養層面的學術承擔，促使教會與整個社會在新興科技所形塑的歷史處境中，重新反思並界定「作為一個人而活」的根本意涵。

¹³ Erik Hoel, “A Disproof of Large Language Model Consciousness: The Necessity of Continual Learning for Consciousness,” (19 Jan 2026): <https://web3.arxiv.org/pdf/2512.12802>