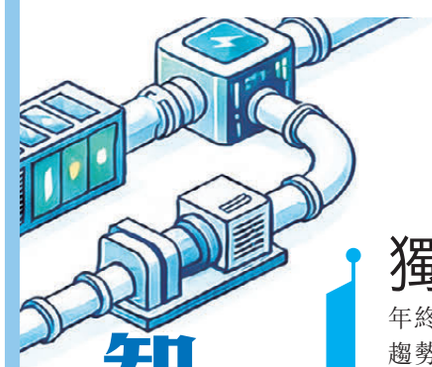
編輯：陳思夢 設計：崔竣明 文：胡藝禾



全球競逐「能源基建」 智能體開啟AI協作新時代

今年初，美國大型科企英偉達（NVIDIA）行政總裁黃仁勳在一篇罕見的長篇署名文章中，將AI產業比作「五層蛋糕」，依次為能源、芯片、基礎設施、模型和應用。他大膽預判，未來幾年傳統的App形態可能消失，一種全新的軟件範式AI智能體（AI Agent）極可能成為主流。AI將從被動的「工具」進化為主動的「同事」。

獨立AI基準測試機構Artificial Analysis發布的《2025年終AI發展報告》印證了這一趨勢。報告總結的五大行業趨勢裏，「AI智能體起飛」位列核心。2025年11月誕生的OpenClaw如今爆紅，宣告智能體走入大眾視野，而今年智能體的應用範圍將擴展到企業級工作場景。信息諮詢公司Gartner的預測更為具體：到2026年底，40%的企業應用將嵌入AI智能體，而2025年這一比例僅為5%。

算力盡頭是電力

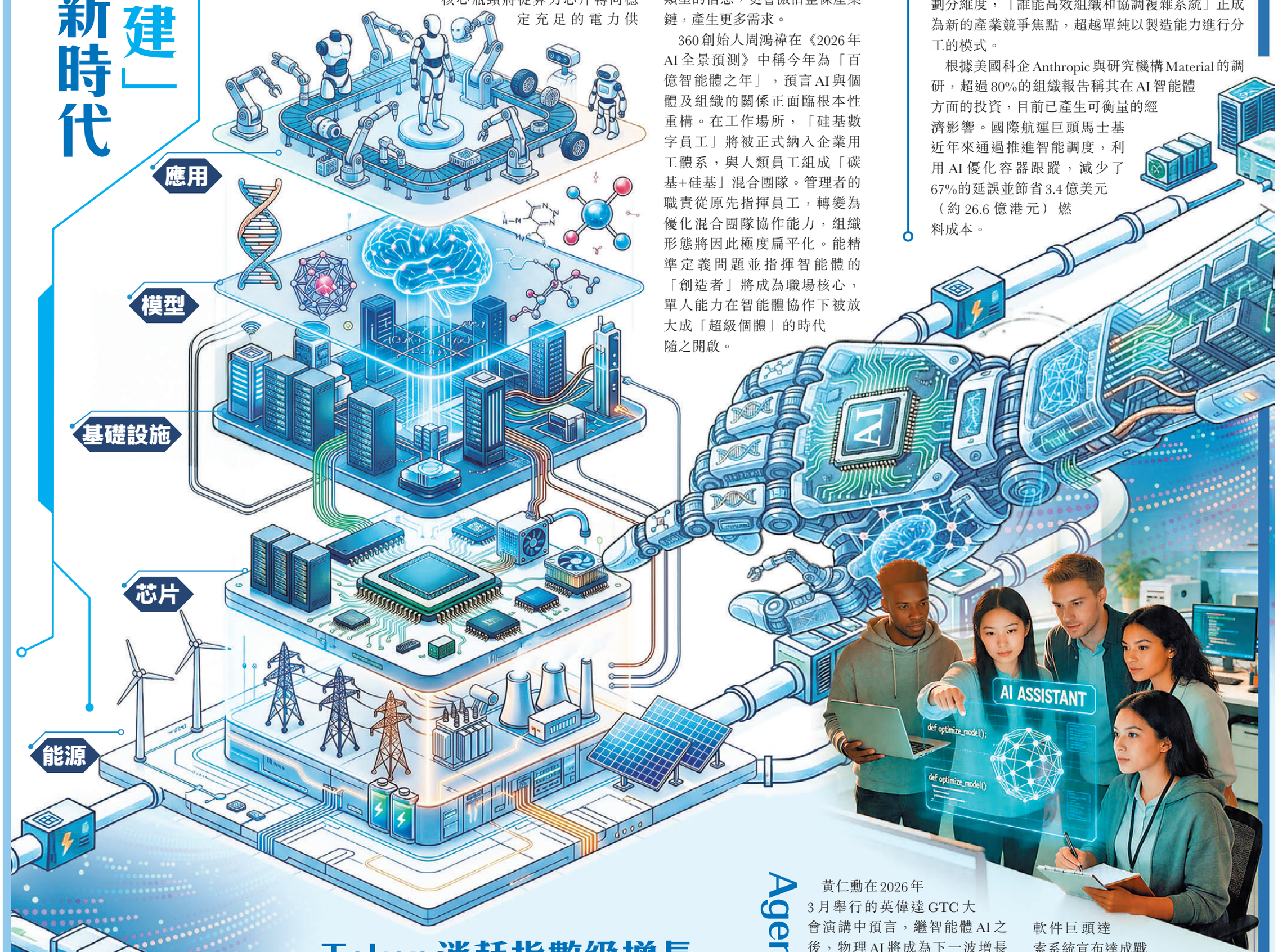
這一爆發式增長背後，是AI產業的發展動力的根本轉變。《報告》進一步指出，2025年推理模型已成行業常態，各大實驗室紛紛推出自家的推理模型，智能水平不斷提升，在通用推理、科學推理、長周期智能體任務以及編碼領域的表現不斷優化。去年推出的推理模型還全面普及工具調用訓練，通過預訓練和強化學習優化，專用於智能體執行任務所需。今年，隨着企業轉向推理應用，將「僱用」更多AI智能體解決實際問題，過去拚「誰家模型更博學」的範式，將變為拚推理算力的大小，因推理模型推動AI智能水平大幅提升的同時，對Token（詞元）數的需求顯著擴大，工作負載規模大增。無論是影視創作還是城市調度，高頻、長流程的推理任務將推動推理算力需求呈現短期指數級增長，重塑能源與芯片的格局。屆時，制約發展的核心瓶頸將從算力芯片轉向穩定充足的電力供給，全球科技競爭將聚焦「能源大戰」。

4月16日，技術人員在中國南方電網公司機器人實驗室內為人形機器人編寫程式。

開源模型激活產業

黃仁勳的「五層蛋糕」理論為理解這一變革提供了宏觀框架。他指出，AI算力正在成為如同電力和互聯網一樣必不可少的基礎設施，每一個成功的應用都會向上拉動其下方的每一層。他特別強調了開源模型的關鍵角色，並以DeepSeek-R1為例，指出當強大的推理模型被廣泛應用時，不僅使AI可以理解更多種類型的信息，更會激活整條產業鏈，產生更多需求。

360創始人周鴻禕在《2026年AI全景預測》中稱今年為「百億智能體之年」，預言AI與個體及組織的關係正面臨根本性重構。在工作場所，「硅基數字員工」將被正式納入企業用工體系，與人類員工組成「碳基+硅基」混合團隊。管理者的職責從原先指揮員工，轉變為優化混合團隊協作能力，組織形態將因此極度扁平化。能精準定義問題並指揮智能體的「創造者」將成為職場核心，單人能力在智能體協作下被放大成「超級個體」的時代隨之開啟。



Token消耗指數級增長 算力通脹浪潮湧起

今年以來，全球雲計算企業相繼釋放漲價信號，亞馬遜AWS、Google雲在北美地區部分服務的價格漲幅達100%。3月18日，阿里雲與百度智能雲同日宣布上調AI算力產品服務價格，漲幅最高達34%。全球雲計算平台的同步漲價潮，將AI算力供需失衡的緊迫性推至台前，揭開智能體時代算力大戰的一角。

算力緊張的背後，是以OpenClaw為代表的AI智能體帶來的Token消耗指數級躍升。與傳統聊天機械人不同，AI智能體具備自主調用工具、長上下文記憶、多工具鏈協同等特點，一次任務動輒消耗數十萬至百萬Token。廣發證券研報指出，Agent應用相比傳統聊天機械人，Token消耗量提升1至2個數量級。科技市場研究公司IDC預計年度Token消耗數量，將從2025年的0.0005 PetaTokens暴增至2030年的152,667 PetaTokens，年複合增長率高達3,418%。

算力需求的爆發正重塑整個產業鏈的利益分配格局。德勤報告指出，在2026年，AI推理任務將佔據全球算力需求約三分之二。英偉達在GTC大會上發布Vera Rubin計算平台，推理性能較上一代提升約35倍，Token生成速率實現350倍增長。行政總裁黃仁勳在演講中明確提出，AI時代的Token是「新貨幣」，智能體AI和物理AI將成為下一階段增長點。

企業需前瞻規劃 應對供需失衡

算力通脹的浪潮已經湧起，此輪漲價並非價格周期波動，而是算力供需失衡下的被動型市場應對。供給側GPU、存儲、帶寬、電力成本剛性上漲，需求側AI訓練和推理需求爆發，資源稀缺性明顯。新市場環境下，平台方如何在高昂成本與用戶留存之間找到平衡，個人用戶又該如何精打細算，都將成為行業發展的新一輪起點。IDC中國高級研究經理孫振亞建議，企業需在算力資源、模型選擇和搭配上作出前瞻規劃，維持可持續的投入產出比。

Google大樓。

智能體滲入供應鏈 重塑全球產業分工

與公眾想像中「面向消費者」不同，當前AI智能體最先滲透的並非終端，而是全球產業鏈中大量高度流程化、規則密集的中間環節。這些環節往往涉及多系統協調、數據整合和實時決策，智能體的介入正在顯著提升效率，並悄然重塑全球產業分工格局。

秒級響應 減少決策延遲

相關實踐顯示，智能體更擅長處理規則複雜、變量眾多的協同問題，而人類管理者則轉向負責應對異常情況及在策略層面進行干預。智能體的應用並不直接改變生產環節，卻悄然重塑企業內部的組織方式和決策節奏。工業巨頭正嘗試將智能體嵌入排產和供應鏈調度系統，用於在多工廠、多供應商約束條件下進行實時協調。西門子的NX CAM Copilot作為AI驅動的編程助手，能優化從設計到生產的全流程，有效突破編程瓶頸。德勤的報告亦顯示，使用智能體的企業在供應鏈決策環節的延遲，從數天降低到秒級。

協同效率成關鍵

Gartner預測，到2026年底，40%的企業應用將集成專屬智能體，這將從根本上顛覆供應鏈運作。上一輪全球化主要圍繞生產環節，不同國家在製造、組裝環節形成分工。而在智能體逐步成熟的環境下，AI的能力不再僅限於數據分析，而可深入生產協調，自主優化供應鏈路徑，減少決策延遲，推動企業從被動響應轉向主動預測。產業鏈的協同權和執行權在這一過程中被重構，全球分工形成新的劃分維度，「誰能高效組織和協調複雜系統」正成為新的產業競爭焦點，超越單純以製造能力進行分工的模式。

根據美國科企Anthropic與研究機構Material的調研，超過80%的組織報告稱其在AI智能體方面的投資，目前已產生可衡量的經濟影響。國際航運巨頭馬士基近年來通過推進智能調度，利用AI優化容器跟蹤，減少了67%的延誤並節省3.4億美元（約26.6億港元）燃料成本。

Agent應用前瞻 物理AI崛起

黃仁勳在2026年3月舉行的英偉達GTC大會演講中預言，繼智能體AI之後，物理AI將成為下一波增長浪潮，其在製造業、機械人、自動駕駛等物理領域的應用潛力，甚至可能改變世界。

這一判斷得到了產業界廣泛呼應。圖靈獎得主LeCun的判斷直指核心：大語言模型是條死路，因為它無法真正理解物理世界如何運作。今年2月，宇樹科技通過「Dimensional」項目，將OpenClaw與G1人形機械人集成，首次讓機械人能夠理解空間和時間：它能感知周遭環境並留存時空記憶。

在這場從虛擬走向物理的變革中，OpenClaw扮演着獨特的角色。國盛證券研報指出，OpenClaw填補了從認知到執行的關鍵缺口，將大模型的語言和推理能力，轉化為機械人等智能硬件可執行的物理動作指令，同時融合傳感器數據、空間感知能力，實現「觀察—規劃—行動—反饋」的全閉環。

輸出「數字員工」重構生產力

今年2月，英偉達與法國工業軟件巨頭達索系統宣布達成戰略合作，共同構建「工業世界模型」，一個經過科學驗證、扎根於物理學的AI系統，可作為工程、製造領域的關鍵任務平台。通過AI的協助，過去需反覆製作物理原型才能完成的測試，如今可在這個平台中以極低成本完成大量平行迭代，產品從原材料到組裝的整個供應鏈流程，都可在虛擬世界中重建。

在黃仁勳看來，物理AI的意義或許不亞於當年互聯網將信息流通的成本壓至接近零。如果說那場革命重塑了信息的生產與分配，這場革命重塑的則是物理世界本身。未來3至5年，AI硬件將不再是人類操作的「輔助工具」，而是直接演化為具有自主任務理解、環境變化感知、靈活執行閉環能力的「數字員工」。硬件產業的爆發不再依賴於賣機器的硬件利潤，而是轉變為直接向社會提供AI勞動力供給，這將深刻改變經濟學的生產要素構成。