



▲黃仁勳介紹Vera Rubin平台。Rubin架構的問世，不僅是硬件技術的代際躍升，更是AI算力經濟的範式革命。

經濟觀察家

本月舉行的拉斯維加斯CES展會上，英偉達創始人黃仁勳手持Vera Rubin平台的身影，標誌着AI基礎設施迎來歷史性轉折。Rubin是最新一代算力平台架構，不僅是硬件技術的集大成者，更構建了「數據中心即計算單元」的全新生態。

# 技術迭代加快 AI 產業提速



Rubin 架構是以美國天文學家薇拉·魯賓命名的，其並非單純的芯片迭代，而是一套整合六款定製芯片的全棧超級計算解決方案。在AI從生成式向代理式、物理式跨越的關鍵節點，Rubin架構通過「極致協同設計」理念，將推理token成本降至前代Blackwell平台的十分之一，MoE（混合專家）模型訓練GPU用量減少四分之三。

## 行業應用陸續落地

在雲計算領域，Rubin架構正在重塑AI即服務（AlaaS）的商業模式。AWS、谷歌雲等服務商借助Rubin的高算力密度與低運營成本，能夠提供更具性價比的AI訓練與推理服務。例如，谷歌雲為Rubin硬件打造了最優運行環境，客戶可通過其平台快速獲取大規模算力，毋須投入巨額資金建設自有數據中心。這種模式的普及，將使AI算力像水電一樣成為普惠性資源。

自動駕駛領域是Rubin架構的重要應用場景。代理式AI需要處理複雜的交通環境數據，進行多步推理與實時決策，對算力的穩定性和能效比提出了極高要求。Rubin平台的低延遲、高可靠特性，使其能夠支持自動駕駛車輛的實時感知與決策系統，結合Alpamayo模型的因果推理能力，大幅提升車輛應對複雜路況的能力。多家自動駕駛企業已宣布採用Rubin平台構建其車載計算系統，推動自動駕駛技術從L4向L5級別跨越。

醫療健康領域的應用則彰顯了Rubin架構的社會價值。在藥物研發中，研究人員可借助Rubin的超高算力模擬藥物分子與靶點蛋白的相互作用，將傳統需要數年的研發周期縮短至數月；在臨床診斷中，AI模型能夠基於醫學影像進行精準分析，輔助醫生提高診斷準確率。Rubin的機密計算功能還確保醫療數據的安全，為AI在醫療領域的合規應用提供了保障。

工業製造領域，Rubin架構推動數字孿生技術實現質的飛躍。通過整合傳感器數據與AI模型，製造企業能夠構建與物理世界實時同步的數字孿生系統，不僅可實現生產流程的可視化監控，更能通過AI的自主優化能力提升生產效率、降低能耗。某汽車製造商採用Rubin平台後，其數字孿生系統的仿真精度提升40%，生產線上的設備故障預測準確率提高至95%，每年節省維護成本超千萬美元。

## 供應鏈瓶頸成制約

儘管Rubin架構展現出強大的技術優勢，但在規模化落地過程中仍面臨多重挑戰。技術層面，CPO（光電共封裝）、浸沒式液冷等前沿技術尚未完全成熟，大規模量產可能面臨良率不足、成本偏高的問題。例如，CPO技術需要將光引擎與芯片緊密集成，對封裝工藝的精度要求極高，目前行業整體良率仍低於80%，制約了其商業化進度。

供應鏈瓶頸成為短期最主要的制約因素。Rubin架構對高端材料的需求激增，導致HVL P4銅箔、Q布（石英布）、M9樹脂等關鍵材料出現供

Vera Rubin在各大領域應用	
雲計算	AWS、谷歌雲等服務商可借助Rubin的高算力密度與低運營成本，提供更具性價比的AI訓練與推理服務
自動駕駛	Rubin平台的低延遲、高可靠特性，能夠支持自動駕駛車輛的實時感知與決策系統，提升車輛應對複雜路況的能力
醫療健康	研究人員可借助Rubin的超高算力模擬藥物分子與靶點蛋白的相互作用，研發周期縮短至數月；在臨床診斷，AI模型能夠輔助醫生提高診斷準確率
工業製造	Rubin架構推動數字孿生技術實現質的飛躍。通過整合傳感器數據與AI模型，製造企業能夠構建與物理世界實時同步的數字孿生系統，除實現生產流程的可視化監控，更能通過AI的自主優化能力提升生產效率、降低能耗

應短缺。據測算，2026年第二季度起，HVL P4銅箔的月度缺口將達500至600噸，而Q布的供應也難以滿足英偉達每月200萬米的採購需求。這些核心材料主要由日韓企業主導，中國廠商雖在加速研發，但短期內仍面臨專利壁壘與產能爬坡的雙重壓力。

國際市場的競爭與地緣政治風險不容忽視。在光模塊、PCB等領域，Finisar、TTM等行業巨頭憑藉技術積累與產能優勢佔據主導地位，中國廠商需要在技術創新與成本控制上實現突破才能獲市場份額。同時，全球科技產業的地緣政治博弈或影響芯片製造、材料供應等關鍵環節，給Rubin架構的全球部署帶來不確定性。

系統兼容性與遷移成本也是客戶面臨的實際問題。儘管Rubin架構兼容Blackwell平台的優化代碼，但對於採用更早期架構或其他廠商硬件的客戶而言，遷移至Rubin平台仍需要進行大量的適配工作。部分行業客戶的legacy系統與Rubin的先進技術存在兼容性差距，如何在保障業務連續性的前提下實現技術升級，成為考驗生態合作夥伴的重要課題。

## 技術演進未來方向

英偉達已經規劃了Rubin架構的長期演進路線，持續推動算力與效能的雙重突破。2027年推出的Rubin Ultra NVL576平台，將會把FP4推理性能提升至15 ExaFLOPS，較當前GB300平台提升近14倍，進一步鞏固在超大規模AI計算領域的領先地位。該平台將採用正交背板設計，單櫃PCB總層數有望突破150層，實現更高的算力密度。

量子計算與AI的融合將成為Rubin架構重要演進方向。NVQLink量子互連技術的引入，為GPU與QPU的協同計算奠定了基礎，未來這一技術將進一步優化，實現量子比特與AI張量計算的高效協同。這種融合架構有望突破經典計算的物理極限，為解決量子化學、材料科學等領域的複雜問題提供全新方案。

能效優化將持續成為技術演進的核心焦點。英偉達計劃通過先進封裝技術、新型散熱材料與智能功耗管理算法的結合，進一步提升Rubin架構的能效比。市場預計到2028年，Rubin架構的每瓦算力將較當前版本再提升2倍。

軟件定義的算力調度將成為釋放硬件潛能的關鍵。未來英偉達將進一步強化Omniiverse DSX AI工廠藍圖的能力，通過AI驅動的動態資源調度算法，實現計算、存儲、網絡資源的

最優配置。針對不同行業的應用場景，將推出更多定製化軟件棧，使Rubin架構能夠快速適配垂直領域的特殊需求。

Rubin架構的普及將推動AI從「生成式」向「代理式」的全面轉型。隨著推理成本的大幅降低與長上下文處理能力的提升，AI將具備自主規劃、多步推理與工具使用的能力，從單純的內容生成工具升級為能夠解決複雜問題的智能體。這種轉型將催生全新的應用場景——從個人AI助手能夠自主完成工作任務，到工業AI系統能夠自主優化生產流程，AI將深度融入社會經濟的各個層面。

算力普惠將激發創新活力，打破科技巨頭的壟斷格局。此前需要數千萬美元投入才能啟動的AI項目，現在可能僅需數百萬美元即可落地，這為中小企業與科研機構提供了平等的創新機會。市場預計未來五年，AI領域的創業公司數量將增長3倍，邊緣創新將成為推動AI發展的重要力量，加速技術迭代與應用普及。

## AI倫理與安全成挑戰

全球計算基礎設施的AI化轉型將進入加速期。Rubin架構推動數據中心從通用計算設施向專用AI工廠轉型，這種轉型不僅體現在硬件設備的更新，更涉及運營模式、能源結構與人才需求的全面變革。市場預計到2030年，全球AI數據中心的算力規模將達到當前的50倍，成為數字經濟的核心基礎設施，推動十萬億美元級別的產業升級。

AI倫理與安全將面臨新的挑戰與機遇。Rubin架構的第三代機密計算技術，通過在CPU、GPU和NVLink域之間創建統一的信任執行環境，為AI模型與數據提供了全方位保護。這種技術保障將促進AI在金融、醫療等敏感領域的應用，但同時也對AI倫理規範與監管框架提出了更高要求。技術創新與倫理規範的協同發展，將成為AI產業可持續發展的關鍵。

英偉達Rubin架構的問世，不僅是一次硬件技術的代際躍升，更是AI算力經濟的範式革命。通過「六芯合一」的協同設計理念，該架構打破了傳統服務器的性能瓶頸，實現了算力、能效與成本的重重突破。

儘管面臨供應瓶頸、技術成熟度等短期挑戰，但Rubin架構的長期發展趨勢不可逆轉。隨着技術的持續演進與生態的不斷完善，AI算力將變得更加普惠、高效與安全，推動人類社會進入智能新時代。

（作者為外資投資基金董事總經理）

# 國家發展AI 港可當超級研發中心



創科瞭望

1月7日，中國工業和信息化部等八部門聯合發布《「人工智能+製造」專項行動實施意見》（下稱《實施意見》）。這份文件不僅是一項產業政策，更是國家推動「新質生產力」、實現高質量發展的重大戰略部署。其核心目標是將人工智能（AI）與製造業進行全方位、深層次、高水平的融合，重塑生產模式與經濟形態。

《實施意見》的發布，標誌着國家正式將AI視為賦能新型工業化的核心引擎，這對正全力推進「新型工業化」的香港而言，既帶來前所未有的歷史機遇，也提出了需要審慎應對的挑戰。

《實施意見》為中國AI與製造業的融合設定了清晰的量化目標。到2027年，旨在實現人工智能關鍵核心技術的安全可靠供給，使產業規模和賦能水平穩居世界前列。具體指標包括：1）模型與數據：推動3至5個通用大模型在製造業深度應用，形成特色化、全覆蓋的行業大模型，打造100個工業領域高質量數據集；2）應用與生態：推廣500個典型應用場景，培育2至3家具有全球影響力的生態主導型企業，並選出1000家標桿企業。

為實現這些目標，政策從供給與應用兩端進行了系統布局。在供給側，強調強化AI算力，支持高端訓練晶片、端側推理晶片等關鍵核心技術的突破，並開發適應製造業特點的高水平行業模型。在應用側，則聚焦於全流程的智能化升級，涵蓋研發設計、中試驗證、生產製造、營銷服務及運營管理五大環節。特別是在智能裝備領域，明確提出要推動機器人、人形機器人等技術的應用，並在數據治理和安全保障方面建立完善體系，顯示了國家在追求技術創新的同時，對安全可控的高度重視。

## 發揮港頂尖科研能力

這一戰略為香港的「新型工業化」進程注入了動力。我們應善用獨特優勢，主動對接國家規劃，將在多個層面迎來重大發展機遇。首先，粵港澳大灣區協同發展將成為香港先進製造業的核心引擎。香港擁有世界頂尖的基礎科研能力和國際化的人才網絡，但在土地成本和產業規模上存在局限。《實施意見》鼓勵的應用場景，如新能源汽車、智能消費電子、生物醫藥等，正是大灣區的優勢產業。香港可以將自身定位為「超級研發中心」，專注於AI晶片設計、底層算法創新、生物科技等高附加值環節，而將生產製造、場景測試等環節與深圳、東莞等大灣區內地城市深度

協同，形成「香港研發、灣區製造」的高效合作模式。

其次，為香港創科生態帶來廣闊市場與發展空間。《實施意見》提出的500個應用場景和1000家標桿企業，構成一個龐大的「藍海市場」。香港的AI初創企業和科技公司，可以針對這些具體場景開發定製化的解決方案。例如，香港在金融科技（FinTech）和合規科技（RegTech）領域積累豐富經驗，可以為內地製造業企業提供供應鏈金融、智能合規等方面的AI解決方案。此外，香港作為國際金融中心，可以為尋求全球擴張的「生態領軍企業」和「標桿企業」提供上市融資、併購重組等專業服務，進一步鞏固金融樞紐地位。

## 加強本地培訓 引進海外專才

儘管機遇巨大，香港在過程中同時面臨一些潛在的挑戰。其一，是技術依賴與供應鏈安全問題。雖然國家力求核心技術自主可控，但在部分高端AI晶片、開發工具等領域，全球供應鏈仍然高度集中。香港作為自由港，在引進國際先進技術方面具備優勢，但也有地緣政治而起的供應鏈風險。其二，是與內地城市的競爭壓力。深圳、上海等城市在AI產業政策、資金投入和市場規模上均具備強大實力，香港必須找準自身不可替代的獨特價值。

筆者認為《實施意見》為香港AI發展帶來深刻啟示。啟示一，人才是第一資源。國家戰略將極大激發對AI與製造業複合型人才的需求，香港必須加強本地人才培養，並積極引進全球頂尖的AI科學家 and 工程師。啟示二，數據是核心要素。香港應積極探索建立安全合規的數據跨境流動機制，使我們在處理國際數據、進行跨境模型訓練方面具備獨特的信任優勢；同時成為連接內地數據資源與國際AI研發的「數據樞紐」。

香港亦應繼續發揮「超級聯繫人」的角色，加強與全球頂尖科研機構和科技企業的合作，將國際前沿技術和管理經驗引進國家製造業體系，同時幫助內地AI企業「走出去」。

總體而言，國家《實施意見》為香港的未來發展指明了清晰方向。香港應該主動思考如何成為國家戰略的「超級增值器」。通過聚焦基礎科研、發揮金融與法律制度優勢、深化大灣區協同，香港完全有能力在國家建設製造強國的宏偉藍圖中，扮演著不可或缺的關鍵角色。這不僅是香港實現自身經濟轉型和「新型工業化」的歷史機遇，更是我們履行國家使命、貢獻民族復興的應有之義。

（作者為香港創科發展協會創會主席）

# 買家看好港前景 工商舖成交旺



樓語縱橫

總結來說，過去一年樓市整體成交絕對是驚喜的。樓市全面撤辣，加上減息周期的重啟，令樓市整體成交量顯著回升。在剛過去的12月份，樓市整體註冊量高達9000宗，創近20個月新高，除了一手交投持續暢旺，二手成交亦明顯回升，連帶近年交投相對淡靜的工商舖市場，上月亦接連出現大額商廈成交。

去年全年樓市註冊量超越8萬宗，為2021年以來的4年新高。新盤全年成交超越2萬伙，是自2004年以來的21年新高，按年急升約28%，二手註冊量亦重上4.6萬宗，按年升幅亦達16%。最令人驚喜的是工商舖市場，去年接連錄得多宗大額成交，令工商舖註冊量顯著增加23%，帶動成交金額回升13%，當中包括3宗高達30億元以上的大額商廈交易。去年成交金額最高是銅鑼灣港島壹號中心的72億元成交，其次是港交所所以63億元購入中環交易廣場1期的9層樓面，上月再錄得京東集團以近35億元購入中環中國建設銀行大廈的50%權益。



▲去年樓市全面撤辣，加上減息周期重啟，令全年樓市成交量顯著回升。