

阿里發布真武 V900 芯片 國產算力最強

明年第一季量產 性能達上一代3倍

阿里巴巴（9988）首席執行官吳泳銘昨在杭州舉行的2026雲棲大會上公布阿里涵蓋AI模型、AI芯片和AI雲的AI三大基礎設施布局的最新線路圖，並發布旗下「平頭哥」研發的下一代AI芯片——真武V900，稱其為「國產算力性能最強」，性能達上一代真武M890的3倍，計劃於明年第一季量產。與此同時，阿里Qwen（千問）團隊正持續推進模型架構和數據優化的研究，計劃訓練5萬億至10萬億參數規模的全新模型，目標完成更複雜、更長程的任務，向超級人工智能邁進。

●香港文匯報記者 周曉菁

「機器正在成為思考的主力，智能正在成為一種規模化的商品，未來機器思考總量將達到人類的1,000倍以上。上一次這樣的變化，是工業革命把動力變成了規模化的商品。」吳泳銘在演講中指出，未來機器智能時代的三大基石是AI模型、AI芯片和AI雲，這是機器思考能夠規模化供給的前提。阿里巴巴將堅定投入這三項基礎設施建設，這是阿里的長期戰略選擇。

將推出Qwen 4.5及Qwen 5模型

在AI模型方面，吳泳銘表示，目前Qwen團隊正在進行RSI（遞歸自我改進）方面的探索，並取得一定的進展。RSI即模型從真實反饋中發現自己的短板，自己設計實驗、構建數據、評價結果，循環推動模型自身的進化。

阿里還表示，其下一代模型Qwen 4現正進行訓練，即將推出的Qwen 4.5及Qwen 5模型，預計參數規模將擴展至5萬億至10萬億；而其目前旗艦模型Qwen 3.8 Max總參數規模為2.4萬億。此外，阿里推出多項多模態模型更新，範疇涵蓋語音、音頻、及視覺模型，如同聲傳譯模型Qwen3.8-LiveTranslate，進一步提升翻譯的忠實度、流暢度及精簡度，並將字均延遲（LAAL）由2.8秒下降20%至2.3秒。

真武AI芯片已服務客戶逾650家

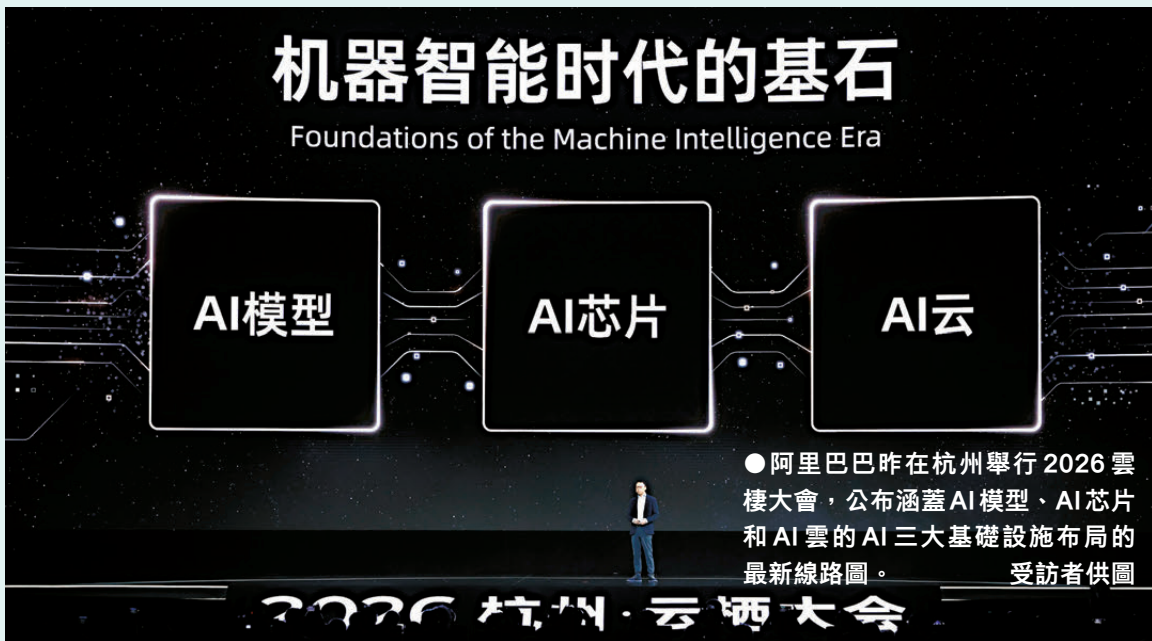
在AI芯片方面，吳泳銘形容，「Token是AI時代的電，芯片就是發電機」。「平頭哥」正在對數據中心芯片進行全面布局——「真武」系列是GPU芯片，「倚天」系列是CPU芯片，「磐脈」系列是智能網卡，還有ICN互聯芯片。組建超大規模AI集群的核心芯片已全面覆蓋。目前，「平頭哥」的真武AI芯片已服務逾650家客戶，涵蓋汽車、金融、大語言模型、具身智能、能源及製造等行業。

吳泳銘表示，阿里自研的M890 AI超節點，已經成功支撐超過2萬億參數大模型的高效推理，是全球極少數具備這樣能力的公司之一。本季度開始，阿里雲已經規模化上架AI超節點。

對於新一代AI芯片真武V900，吳泳銘在演講中表示，這是目前中國算力性能最強的AI芯片，性能達到M890的3倍，基於它構建的單一集群可擴展至50萬卡，支撐前沿模型的訓練和推理。

吳泳銘還說，由於平頭哥芯片產品線的成熟和客戶的廣泛應用，預計平頭哥AI芯片的年出貨量將大幅提升。

AI雲方面，吳泳銘指出，目前客戶的AI需求非常強勁，中長期需求遠超供給能力，阿里將與所有合作夥伴一起，全力投入AI基礎設施建設，目標是到2032年，阿里雲運營的全球數據中心規模超過20GW。



●吳泳銘在2026雲棲大會上表示，阿里雲將全面布局AI基礎設施，包括AI模型、AI芯片和AI雲。受訪者供圖



●觀眾在杭州國際博覽中心智能館基礎大模型展區參觀。新華社

●AI模型：Qwen團隊計劃訓練5萬億至10萬億參數規模的全新模型，向超級人工智能邁進

●AI芯片：平頭哥發布AI芯片真武V900，算力提升至真武M890的3倍，基於它構建的單一集群可擴展至50萬卡，可支撐前沿模型的訓練和推理，計劃於明年第一季量產

●AI雲：到2032年，阿里雲運營的全球數據中心規模將超過20GW

整理：香港文匯報記者 周曉菁

蔡崇信：阿里堅定布局五層全棧AI體系

香港文匯報訊（記者 周曉菁）阿里巴巴董事會主席蔡崇信昨在雲棲大會上表示，阿里巴巴正在堅定布局五層全棧AI體系，繼續打造從芯片、雲、模型到模型服務和智能體應用全棧體系，這五層的體系彼此協同，構成從技術能力到產業應用的完整路徑，讓每一項技術進步都能更快轉化成產業進步。

他稱，今年雲棲大會的主題是「智以致用」。過去幾年，AI不斷拓展人們對技術能力的想像。AI有着廣闊的發展空間，能在真實場景落地以及規模化應用，並能提升千行百業的生產力，讓AI

從技術突破走向價值創造。面對全球市場的不確定性，蔡崇信提出，越是複雜環境，越要堅持開放合作與務實創新。科技不會自動產生價值，只有扎根產業、聯合夥伴，才能釋放AI生產力。

下一代自研CPU2027年推出

在大會現場，阿里巴巴同時公布面向AI智能體任務的下一代自研CPU路線圖，相關產品計劃於2027年推出。相較倚天710，倚天720具備更強的單核效能、更高的核心密度及更佳能源效率；

倚天730則是首款採用平頭哥自研微架構的CPU，SPECint2017/GHz效能較倚天710提升最多40%。

此外，阿里雲圍繞Agentic Cloud戰略推出一系列全面升級，聚焦Model、Harness及Context三大核心場景，並構建三個相應層級：面向模型的AI Native Cloud，支持大規模模型訓練及推理；面向Agent運行框架的Agent Native Cloud，支持企業級智能體部署、運行及安全；以及面向上下文的Context Engine，提供實時數據及長期記憶，協助客戶大規模運用AI創造實際價值。

廣州年產60億顆5G/6G濾波芯片基地投產

香港文匯報訊（記者 敖敏輝 廣州報道）歷經20多年攻關，中國研發人員突破美國巨頭數十年的技術壟斷，成功研發出性能更優、成本更低的手機通信關鍵元器件濾波芯片。該通信領域「中國芯」在粵港澳大灣區迎來大規模量產關鍵節點，廣州市艾佛光通科技有限公司（下稱「艾佛光通」）SABARR總部基地昨日在中新廣州知識城舉行產線點亮儀式，這一佔地50畝的生產基地，可年產60億顆5G/6G SABARR（基於單晶氮化鋁的體聲波濾波技術）濾波芯片，提高國產高頻濾波芯片的自主供應能力。根據規劃，至2030年，該基地濾波芯片產能將擴充至150億顆以上。

安裝在手機射頻前端的體聲波濾波芯片（BAW濾波器），被稱為智能手機「信號守門員」，比米粒還小，卻決定手機能否穩定接收信號、無線網絡是否流暢，是5G手機、基站等無線通信設備不可或缺的關鍵器件。長期以來，BAW濾波器市場幾乎被美國兩家企業壟斷，通過知識產權壁壘佔據全球95%以上市場份額，導致中國在通信產業鏈關鍵環節面臨「卡脖子」問題。

當前，主流體聲波濾波器多採用多晶氮化鋁材料，以華南理工大學教授李國強領銜的團隊則另闢蹊徑，開發出基於單晶氮化鋁的體聲波濾波技術（註冊商標SABARR），並創立艾佛光通。

艾佛光通副總經理、研發負責人衣新燕告訴記者，技術突破和工藝製程的改進，帶來國產濾波芯片在性能及成本方面均大幅改善。由於單晶材料是整齊排列的，信號傳輸又快又準，而多晶材料就像散亂的隊伍，容易產生能量損耗。因此，



●艾佛光通 SABARR總部基地昨日在中新廣州知識城舉行產線點亮儀式。香港文匯報記者敖敏輝 攝

SABARR濾波芯片核心性能比國外產品提升30%，並通過芯片架構的重新設計，成本相較同類下降40%。

帶動上下游材料封測等灣區集聚

此前，艾佛光通在廣東河源建設工廠，近三年累計出貨超20億顆芯片，應用於包括華為、中興、三星、大疆等品牌手機、基站、無人機等領域，以及雷達、航空航天、火星探測等重大工程領域。

此次艾佛光通新總部基地投產，將把產能推向新量級。衣新燕介紹，此次開通的產線為一期工

程，預計實現年產60億顆芯片；後續基地將繼續擴產，遠期目標達到150億至200億顆。

「關鍵核心技術是要不來、買不來、討不來的。」艾佛光通創始人、總經理李國強說，總部基地的啟用將帶動上下游材料、設備、封測等配套資源在大灣區集聚，對補齊中國高端射頻芯片產業鏈短板、助力廣東打造全國半導體與集成電路「第三極」具有重要意義。

目前艾佛光通已實現近100款濾波芯片的穩定量產。至於下一步研發方向，衣新燕表示，基於現有的產品，將不斷向低成本、小型化去發展。同時，還將布局高集成度的芯片模組等。

測評：小米新MiMo大模型奪全球開源國產雙第一

香港文匯報訊（記者 周曉菁）小米（1810）昨正式發布並開源全新一代MiMo大模型，包括全模態旗艦模型Xiaomi MiMo-V2.6-Pro、高效推理模型Xiaomi MiMo-V2.6-Flash、MiMo-V2.6系列維持API定價不變。Artificial Analysis測評指，Pro版本登頂全球開源大模型第一、國產模型第一，成為全球開源模型陣容中唯一的全模態Pro模型。

數據顯示，MiMo-V2.6-Pro實測單任務成本為0.13美元，在同等智能水平下，價格僅為海外部分模型的1/20至1/60。此舉首次將全球前沿智能帶入0.1美元成本區間，大幅推進全球「智能×成本」帕累托前沿。

測評稱，經過大規模、多任務強化學習（RL）後訓練，MiMo-V2.6-Pro在綜合智能指數中取得優異表現，僅次於GPT-6Astra、Claude opus5等頂級開源模型。

實現智能能力跨代躍升

MiMo-V2.6系列預訓練架構和參數規模與上一代保持不變，但實現了智能的代際躍升，在Coding、Agent、Design、Research、Cyber能力上全面增強。

最顯著的突破在於MiMo-V2.6-Pro開始以「Co-Scientist」的角色參與真實科研。在小米前沿

材料研究團隊的交互下，MiMo-V2.6-Pro協助開展材料設計與計算篩選，提出數種金屬有機框架（MOF）材料方案，用於吸附全氟和多氟烴基物質（PFAS）。模型不僅完成文獻與專利梳理，還調用開源計算工具搭建模擬環境，協助團隊將研發周期從約一個月縮短至2至3天。

小米昨日亦同步上線MiMo-V2.6-Pro的超高速模式V2.6-Pro-Ultraspeed，以及MiMo Desktop桌面客戶端正式版本。

小米還宣布，將於9月23日（星期三）晚上7時舉行小米秋季新品發布會，全新小米18 Pro系列將於發布會上正式亮相。

掃地機器人零部件商歡創科技招股

香港文匯報訊（記者 莊程敏）新股市場暢旺，再有前沿科技股來港IPO。主要向掃地機器人供應空間感知零部件的歡創科技（6802）昨日起招股，擬發行1,158.88萬股H股，其中10%股份作公開發售，招股價58.85元，集資約6.82億元，每手100股，入場費5,944.35元，預期於9月30日掛牌，中金公司、國信證券香港為聯席保薦人。

該公司聯合創始人周琨昨表示，公司2026年第一季實現了扭虧為盈，淨利潤達到776萬元人民幣，完成了關鍵性的轉折，主要因為新產品銷量增加和其毛利率較高，此次高毛利新產品顯著改善了公司的盈利能力，目前這一增長態勢仍在延續。

周琨續指，「自2018年首款芯片流片以來，我們已迭代開發三代芯片，累計流片量超過3,000萬片。正是基於這項技術投入與構築的護城河，使公司得以在產業中實現超越，穩居行業龍頭地位。」

另有4隻前沿科技股招股

另外，日前亦有4隻前沿科技股齊招股，合共吸資143.6億元，包括特專科技公司（18C）本末科技（6731）、羅博特科（3757）、景旺電子（3228）和彤程新材（9607），分別集資10.8億元、51.8億元、51億元和30億元。4隻新股都將於本周四截止招股，本月29日掛牌。

以基石投資者來看，集資最多的羅博特科引入最多基投，共16個，包括淡馬錫、劍橋科技（6166）、A股天孚通信聯合創始人朱國棟等，總認購額佔是次最高集資額51.8億元的35.2%。印製電路板（PCB）產品製造商景旺電子亦有14名基投，包括中際旭創（3308）、大族數控（3200）、劍橋科技、建滔集團（0148）等，總認購額佔是次最高集資額51億元的47.7%，而大族數控更是使用外部融資來認購股份。



●歡創科技聯合創始人周琨（右）。香港文匯報記者 莊程敏 攝