

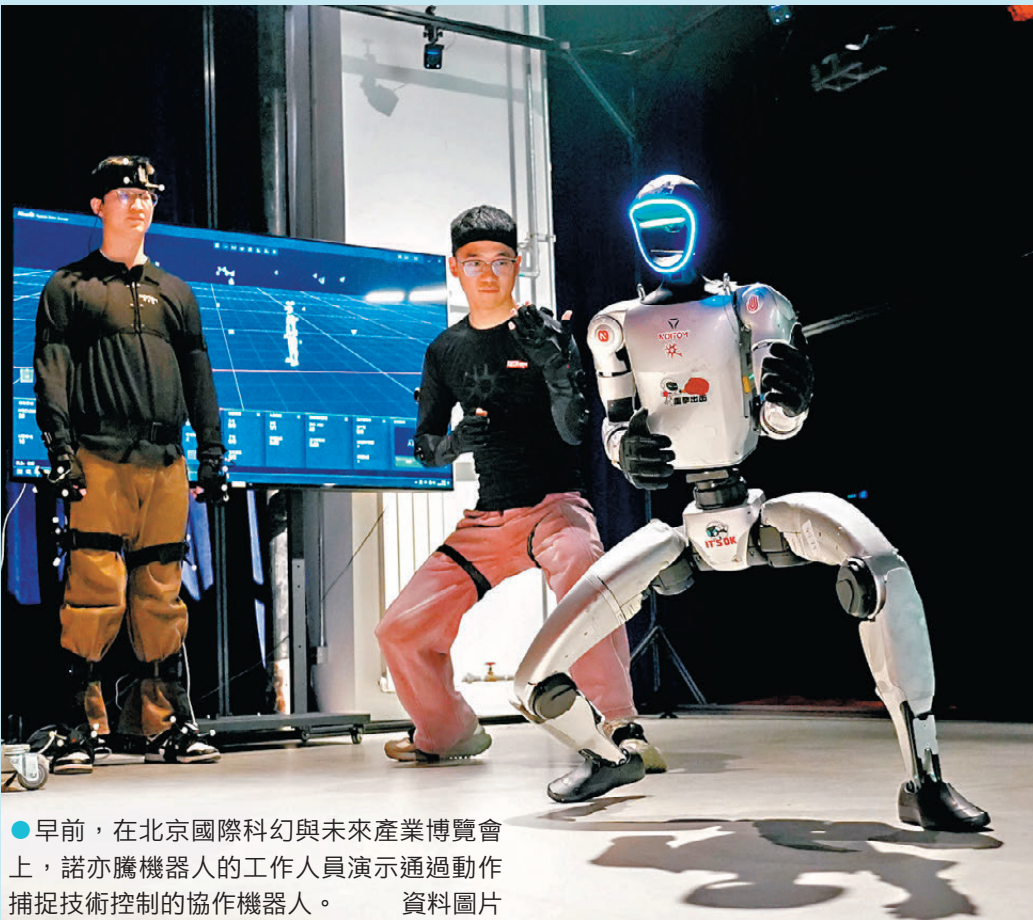
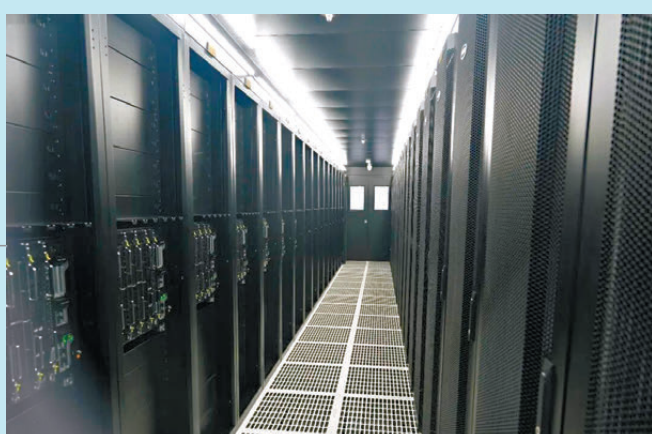
北京首家詞元工廠落地

適配各類複雜AI場景



▲北京首家Token工廠——北京壹號詞元工廠9日正式落地亦莊。香港文匯報北京傳真

▶北京壹號詞元工廠由一台台服務器組成。香港文匯報北京傳真



●早前，在北京國際科幻與未來產業博覽會上，諾亦騰機器人的工作人員演示通過動作捕捉技術控制的協作機器人。資料圖片

一期日產能達1.4萬億 九成任務響應時間低於10秒

北京建設「人工智能之城」再落一子，北京首家Token（即詞元，AI處理文字最小單位）工廠——北京壹號詞元工廠9日正式落地亦莊，並同步向全球開源「詞元工廠性能基準」，項目一期每日Token產能可達1.4萬億。詞元工廠對標電網穩定運行標準，能實現7×24小時不間斷運行，一半任務能在6秒內響應，九成任務響應時間低於10秒，全面適配各類複雜AI場景。

●香港文匯報記者 馬曉芳 北京報道

今年3月，中國日均詞元調用量已突破140萬億，兩年增長超千倍。據摩根大通預測，2025年至2030年，中國Token消耗量年複合增長率將高達330%，5年增長400倍。有機構測算，當前中國國產詞元定價約為海外的1/10，中國詞元價格優勢極為明顯。

像用電一樣用Token

Token算力是服務器運算的能力。在人工智能時代，算力已成為國家綜合實力的核心指標。數據顯示，目前國內單日Token消耗量在9個月內暴漲4倍，用量從萬億級躍升至千萬億級。從日常人機對話、專業科研分析，到高清視頻生成，不同場景對算力的要求天差地別，精細化服務成為剛需。算力越強，每秒能處理的Token就越多，使用成本隨之降低。大語言模型正從簡單的對話機器人加速演進為能夠自主規劃、推理並採取行動以達成複雜目標的長時運行系統。一個簡單的用戶指令背後可能是模型數十次推理循環，工具調用與自我反思的疊加。這意味着算力消耗不再是「一問一答」式的短時脈衝，而是長周期、多輪次、高不確定性的混沌負載。

北京壹號詞元工廠是面向智能體時代的「新型電廠」，實現像用電一樣用Token。該工廠由軟通動力信息技術股份有限公司建設，由一台台服務器組成，項目一期每日Token產能可達1.4萬億。工廠遠期目標實現日產10萬億Token，未來還將聯動張家口、烏蘭察布等綠電基地，構建京津冀一體化算力集群，讓算力生產隨着規模持續擴大，大幅降低各行各業使用AI算力的門檻。

大模型算力供給邁流水線時代

軟通動力高級副總裁、軟通智算總裁劉怡表示，科研人員梳理25年以來的數千萬份文獻用來

撰寫學術綜述，僅這一項任務就要消耗1億Token。詞元工廠對標電網穩定運行標準，能實現7×24小時不間斷運行，而且核心響應指標表現優異。一半的任務能在6秒內響應，九成的任務響應時間低於10秒，波動性控制在20%以內，全面適配各類複雜AI場景。詞元工廠不是簡單堆算力，而是把底層資源高效轉化為標準化的「數字燃料」Token，像電網一樣給出明確的服務質量承諾：服務可用性≥99.9%，對標電網供電可靠性≥99.9%；首字延遲P90<10秒、波動<20%，對標電網電壓合格率波動≤±5%；緩存命中率≥90%，對標電網功率因數，減少計算浪費。

當前行業普遍缺少面向智能體場景的算力服務質量評測標準，導致「高分低能、參數虛標、用戶選擇困難」。香港文匯報記者獲悉，伴隨工廠落地，同步向全球開源「詞元工廠性能基準」，實現對算力集群真實服務能力的精準評估與公平對比。這是行業首次針對智能體長時運行特徵建立統一的性能度量標準，大模型算力供給從粗放式吞吐比擠進入標準化、工業化的Token流水線時代。

將推實時監測等動態監控質量

北京亦莊匯聚了大量AI模型企業與科創主體，天地一體化算力網絡持續完善，產業集聚優勢顯著，是此次詞元工廠選址亦莊的重要原因。據了解，目前軟通動力正深度參與平潭兩岸融合算力中心、韶關公共算力服務平台等國家一體化算力網絡重點項目建設，並依託睿動智能體平台，打造連接國際雲廠商與國產AI新勢力的全球化AI雙向樞紐。未來，軟通動力將持續推進「Token工廠」建設，從消費者視角出發推出「實時監測」等一系列質量監控方法，在運行端動態監控Token生成的幻覺率、語義一致性及毫秒級延遲波動。

創新資源得天獨厚 AI企業超兩千家

特稿

北京正大力加強基礎研究，推動科技創新和產業創新深度融合，全力建設「人工智能第一城」，力爭把人工智能和綠色低碳兩個產業培育成新的萬億級產業集群，人工智能產業發展條例也納入了今年北京市的立法計劃。北京市委常委、副市長靳偉日前表示，北京市人工智能企業超2,500家，人工智能核心產業規模約4,500億元人民幣，累計備案上線大模型225款，全國最多。

北京作為全國科技創新中心，擁有得天獨厚的創新資源稟賦。靳偉介紹，北京有近百家高校、千餘家科研機構，全國近一半兩院院士，全國重點實驗室達到145家，佔全國總量的近三成，布局市級重點實驗室401家，在量子信息等前沿領域建設10家新型研發機構。據悉，北京正力爭把人工智能和綠色低碳兩個產業培育成新的萬億級產業集群，推動集成電路、機器人和智能製造、智能網聯汽車等一批千億級產業集群邁向更高水平。

北京市科學技術委員會、中關村科技園區管理委員會主任張繼紅此前透露，北京已經是中國人工智能創新策源地和產業高地。北京人工智能領域頂尖人才佔全國總量的37%；從創新平台來看，北京有人工智能領域全國重點實驗室22家，位居全國首位；從企業看，2025年，培育人工智能獨角獸企業40家，佔全國一半以上。去年，全市人工智能領域投融資405起，累計融資金額552億元，數量和金額都居全國首位。國內人工智能芯片第一股寒武紀、國內大模型第一股智譜華章都誕生在北京。

●香港文匯報記者 馬曉芳 北京報道



●早前，全球首家具身智能機器人4S店Robot Mall在北京亦莊正式開業。圖為藥品分揀機器人吸引參觀者。資料圖片



中國6G將實現網絡無所不達 助力全息互動

香港文匯報訊（記者 郭瀚林 北京報道）國新辦10日舉行中外記者見面會，中國移動研究院未來研究院前沿技術研究員張雨童在會上介紹，中國通信業曾走過一段艱辛的里程，從1G空白、2G跟隨到3G突破、4G並跑，再到現在5G引領世界，這份領先來之不易。未來6G將會做到網絡無所不達、算力無所不在、智能無所不及。

「大家可能都會有過這樣的體驗，在演唱會上明明手機信號滿格，但朋友圈就是發不出去，或者在小區的邊緣以及室內角落視頻畫面會出現卡頓情況。」張雨童舉例，在未來6G規模化落地後，網絡化協作傳輸技術能夠徹底解決這些痛點問題，無論是身處於萬人場館的正中央，還是在遠離基站的邊緣位置，網絡都會主動調度附近多個基站，協同服務，用戶會獲得最穩定最流暢的體驗。此外，在日常生活中，大家可以和遠方的親友實現全息互動，像面對面一樣坐着聊天，讓科幻作品當中的各種想像變成現實。

她談到，在生產領域，6G也將推動工業互聯網的全新升級，以及在社會治理方面，她所在的團隊現在已經在多地建成了通感試驗網，能夠全面覆蓋低空監管以及車路協同等多個場景。未來6G也會全面助力

數字孿生城市建設，對交通、環境以及能源這些基礎設施進行實時監測和智能管理。「總而言之，未來6G將會做到網絡無所不達、算力無所不在、智能無所不及。」

AI為鋼鐵廠降本超千萬元

中國寶武鋼鐵股份設備部機械設備主任工程師馬遵農在會上介紹，人們對傳統鋼鐵工業的印象還停留在工人值守高爐、軋機，在鐵水板坯旁作業，但如今人工智能已開始改變工作方式。他透露，在過去的兩個月，寶鋼一號高爐進

行了原地大修，它所搭載的「寶鋼AI智慧高爐大模型」能夠突破高爐爐內工況無法直接觀測的黑箱問題，通過預測來實現爐內工況的在線，核心模型命中率超90%，一年可降本超過千萬元。另外，截至今年4月，寶鋼已經上線100多個智能體、600多個AI應用場景，其中就連續被譽為「鋼鐵皇冠上的明珠」的取向硅鋼也在AI加持下，板形精度提升了20%，斷帶率下降65%，產品直接運用了在白鶴灘水電站等國家大型工程。



●香港文匯報AI製圖

內地擬三年內網絡初步實現高等級自智

香港文匯報訊 據新華社報道，工業和信息化部10日公布《「人工智能+信息通信」創新發展實施意見（2026—2028年）》，提出到2028年，信息通信網絡初步實現高等級自智，形成30個以上高價值典型場景，打造一批典型應用和特色智能體；城域算力1毫秒時延覆蓋率不低於75%。

人工智能技術正處於快速迭代、加速突破的爆發期，人工智能與信息通信深度融合為新興研究方向，具有較強的複雜性與系統性，在關鍵技術突破、融合路徑探索、商業模式創新等方面仍面臨系列問題和挑戰。

實施意見明確，到2030年，人工智能與信息通信網絡融合關鍵核心技术取得顯著突破，通感算智一體化服務能力大幅提升，形成完備的協同創新和產業生態體系，「人工智能+信息通信」步入技術引領、產業繁榮、安全可靠、智能普惠的發展新階段。

開展AI驅動新型網絡架構研究

實施意見圍繞推動信息通信行業智能化升級，夯實人工智能發展底座、深化融合應用創新推廣、增強信息通信行業治理能力等4個方面部署17項具體任務，包括開展人工智能驅動的網絡架構研究；突破大小模型協同、多智能體協同、智能體通信等技術；加快算力大通道建設，加快構建全國一體化、集約化、市場化的算力服務體系等。

實施意見還以專欄形式部署網絡自智能力提升行動、智算網絡技術產業能力提升行動、智算業務服務能力提升行動。