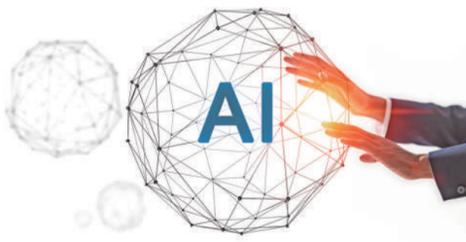


專訪工業富聯首席數據官劉宗長:

AI正從試點走向大規模應用 算力供應明年或將改善



深刻影響,尤其對於公司所在ICT(信息通信技術)行業,是一個非常大的助力。

從公司來看,今年受益於伺服器出貨量大規模增長,公司毛利率、盈利能力、業績都有提升,但這只是開始,這種需求主要來自大模型訓練及數據中心新建,未來人工智能發展,對網絡的需求也很大,將激發大量的網絡交換機的需求,公司在交換機領域也服務很多客戶,如思科等。

另外,大模型也出現了一些小型化趨勢,即實現邊緣端部署,未來這方面會帶動公司較多的企業級伺服器需求,而企業級伺服器也會往AI伺服器方向發展,未來企業需要私有化部署算力時,市場也會看到CPU+GPU混合型算力產品面市。

此外,隨著算力部署越來越多,人工智能應用越來越豐富,未來手機、可穿戴、平板、智慧家庭都會出現非常多的對用戶體驗提升的需求,這會帶動新一輪消費終端需求,所以工業富聯核心業務“雲網端”都將受益,在可以預見的未來3到5年內,都會有一個非常好的增長。

打破技術壟斷

證券時報記者:現在人工智能技術在公司生產管理方面應用情況如何?

劉宗長:三五年前,人工智能在生產、管理環節可能主要是試點,比如一些示范工廠或產線應用了AI技術,它們在技術指標上可能已達到公司需求,但由於一些工程化難題、成本等,難以規模複製,但現在已發展到了一個轉折點,正在慢慢地規模化,這主要得益於技術進步與內部對AI技術接受程度提升。目前,公司開發的AI自動檢測工廠已在產品檢測流程實現規模應用,另外,產線過程中的先進分析、關鍵工藝的参数調整、品質管理、報告生成等場景下,AI技術也已被大範圍使用,而在設計研發環節,公司也在試點讓AI做一些結構化参数設計。

證券時報記者:未來工業富聯國內工廠數字化升級目標是什麼?

劉宗長:怎樣去實現AI的大規模應用,不僅包括AI跟設備的結合,還包括結合後的控制、智能化調參,這些都是未來公司數字化升級的重要方向。

證券時報記者:人工智能對社會有怎樣的具體影響,有沒有一些可以想象的場景?

劉宗長:它會帶來效率的提升,未來隨着人工智能發展,社會會誕生超級個體,即它既是行業專家,又很擅長使用生成式AI工具,這樣1個人的效率可能就是原來10個人的效率,同時也會給未來的就業市場、工作模式帶來很大變化。

另外,生成式AI所引發的人工智能浪潮,還將打破此前大型企業對於AI人才、技術的壟斷格局,降低個人或中小企業創業門檻,他們可以以較低的成本去調用一些AI資源,比如底層大模型,然後生成或研發自身的產品,如垂直大模型,或使用一些生成式AI技術,幫助個人或企業實現更好的發展,這是非常值得期待的方向。

證券時報記者 嚴翠

2023年,以ChatGPT為代表,生成式人工智能(AI)迅猛發展,大有燎原之勢。在此背景下,工業富聯(601138)作為英偉達長期深度合作夥伴,ChatGPT核心算力供應商,受到資本市場的高度關注。日前,證券時報“中國智造面對面”採訪團走進工業富聯,對話公司首席數據官劉宗長,為投資者帶來全球視角下人工智能及算力發展浪潮的最新解讀。

“AI技術正從試點走向大規模應用,明年算力供應會改善,未來算力價格將趨於合理。”劉宗長表示,“人工智能將打破技術壟斷,有助於個人和企業更好發展。”

全產業鏈優勢

證券時報記者:工業富聯和英偉達具體合作情況是怎樣的?

劉宗長:公司2017年開始與英偉達合作開發伺服器產品,當時推出了全世界第一台AI伺服器HGXI,OpenAI創立時,有一張照片展現的是英偉達CEO黃仁勳贈送了一台AI伺服器給馬斯克,並且在上面簽了名,當時那款伺服器就是工業富聯與英偉達共同研發的第一代AI伺服器,所以公司與英偉達合作已久,且非常密切,從產品開發、GPU(圖像處理器)模組和基板的供應、到整個板卡的供應、伺服器組裝,包括供應管理、AI技術賦能、數據訓練和價值再造等,公司是英偉達在全球唯一一個可以覆蓋伺服器全產業鏈的供應商,當然越到產業鏈後端,毛利率越低,如組裝方面,但公司與英偉達的合作是ODM模式,公司深度參與前端產品定制、開發等。

目前,工業富聯不僅是英偉達A100、H100板卡獨家供應商,也是最新的GPU HPC平台獨家設計生產交付供應商,同時,工業富聯今年還首次拿下了英偉達HGXA1伺服器芯片基板訂單,加上早已拿到的DGXA1伺服器芯片基板訂單,工業富聯已成為英偉達AI伺服器芯片基板最大供應商,占據超過50%市場份額。此外,英偉達最新發布的“地表最強”AI芯片GH200的芯片模組訂單等也全部交給了工業富聯。

英偉達還與鴻海集團合作建設人工智能(AI)工廠,這個項目由工業富聯具體推進,未來計劃把AI工廠應用到各個產業。

證券時報記者:公司在算力業務方面核心競爭優勢是什麼?

劉宗長:公司算力業務因為以ODM為主,所以非常考驗技術積累,公司十年前就已布局算力領域,已積累大量研發工程師,以及深厚研發設計、研發製造經驗,且積累了不少優質客戶,包括微軟第一代AI伺服器也是由公司與微軟共同開發的。

另外,在產品製造環節,伺服器是一個高價值產品,一旦製造環節發生問題損失非常大,公司相關產品毛利高低取決於公司良率,如果良率超過客戶要求水平,公司毛利自然會更高,相反就會犧牲掉很多利潤,公司在產品製造方面,具備較強的品質穩定性的把控能力,而且除了幫客戶設計產品,還要幫助客戶設計與之匹配的產線,這些產線全部是自動化、智能化的。

此外,全球供應鏈管理能力也是一大優勢,比如對於一些緊俏貨源的採購能力、庫存管理能力,以及當客戶需要快速產爬坡時公司快速的全球採購能力,包括將產品生產後運到全球所需的物流能力,所以公司在算力方面具備



工業富聯全系列先進散熱AI伺服器

設計、研發、製造、供應鏈管理等多方面優勢,也正因此,今年算力需求爆發時,無論是配合客戶產品開發還是快速量產,公司與同行相比都展現出了非常明顯的優勢。

算力供應將會改善

證券時報記者:公司在算力業務方面有何發展規劃?

劉宗長:首先我們會持续提升公司在算力方面的研發實力,其次公司會更好地服務好現有客戶,幫助它們持續開發新產品,同時,未來AI伺服器散熱將是一大挑戰,所以公司會發力液冷伺服器、綠色數據中心等。

公司還在配合客戶布局新的產能,因為目前來看,英偉達GPU卡價格一直上漲,主要由於產能跟不上,當產能建設逐步投產開始爬坡後,它的出貨量會明顯改善,而且隨着供應恢復,價格也會有所下降,進而會帶動更多新的需求。

此外,在英偉達產品發布會上,黃仁勳曾展示過一段視頻,就是一個熄燈工廠,全部自動化地在生產GPU卡,那個其實是工業富聯的工廠,里面所有自動化設備都由公司自主研發,未來公司仍將繼續拓展相關自動化技術、設備的對外賦能。

證券時報記者:您剛才提到算力的成本會下降,針對這方面能否具體談談?

劉宗長:算力成本高一是因為需求大幅增

長,產能不夠;二是由於地緣政治、美國芯片禁令影響,明年在供應端會有一個明顯提升,同時需求也將在較長時間內保持持續增長,當供應與需求開始平衡後,算力價格會逐漸趨於合理,回歸理性。

核心業務全面受益

證券時報記者:您如何看待當下人工智能產業的發展現狀和未來機遇?公司業務將如何受益於這波人工智能發展熱潮?

劉宗長:OpenAI的ChatGPT火熱後,人工智能迎來了iPhone Moment,會帶來非常大的變化,不光對於人們日常生活,對產業界也將帶來

人工智能風起雲湧 算力基建支撐加速前進

(上接A1版)

“算力基礎建設成為了新基建的重要內容之一,國家統籌建立的‘東數西算’八大樞紐、國家超算中心等,就相當於高速公路的國道,各個地方建立的算力中心等則相當於省道,國家把算力新基建建好了,以後的產業才会有更大的發展機會。”拓維信息董事長李新宇對證券時報記者表示。

國家超算長沙中心主任李青立告訴證券時報記者,儘管“東數西算”等可以部分解決東西部算力需求和資源不匹配的問題,但是還做不到像電力一樣方便地輸送,因為算力輸入是硬件,目前標準等各方面還不統一,技術方面還存在難點。“在這個背景下,當下各城市根據自身算力需求,在本地建設算力中心,也是比較現實且可循環的方式。”李青立說。

國際數據機構IDC聯合其他機構發布的《2021—2022全球計算力指數評估報告》顯示,計算力指數平均每年提高1個點,數字經濟和地區生產總值將分別增長3.5%和1.8%。一個國家或地區增加對算力的投資可以帶來經濟的增長,且這種增長具有長期性和倍增效應。

另據浪潮人工智能研究院測算,預計到2025年全球計算力規模將比2020年提升30倍。

像用電一樣使用算力

近些年,我國數字經濟轉型逐漸深入,

智能化轉型大力推進,算力發展政策東風頻吹。

2022年2月,國家發改委等多部門印發通知,全面啟動“東數西算”工程,同意在京津冀、長三角、粵港澳大灣區、成渝、內蒙古、貴州、甘肅、寧夏等8地啟動建設國家算力樞紐節點,並規劃了10個國家數據中心集羣。

今年以來,從國家到地方的各種算力重構措施、方案也頻出。10月8日,工信部等發布《算力基礎設施高質量發展行動計劃》,提出要着力推動算力基礎設施高質量發展,隨後,上海、廣東、深圳、成都等多地相繼出臺算力發展政策。

12月26日,國家算力政策再加碼。國家發改委、國家數據局等五部門聯合印發《深入實施“東數西算”工程 加快構建全國一體化算力網的實施意見》(下稱《意見》),提出到2025年底綜合算力基礎設施體系初步成型等一系列目標。

《意見》旨在以算力高質量發展賦能經濟高質量發展為主線,協同推進“東數西算”工程,形成跨地域、跨部門發展合力,助力網絡強國、數字中國建設,提出到2025年,我國綜合算力基礎設施體系初步成型,用戶使用各類算力的易用性明顯提高、成本明顯降低,國家樞紐節點間網絡傳輸費用大幅降低;算力網關鍵核心技术基本實現安全可靠,以網絡化、普惠

化、綠色化為特征的算力網高質量發展格局逐步形成,同時強調,將從通用算力、智能算力、超級算力一體化布局等五個統籌出發,推動建設聯網調度、普惠易用、綠色安全的全國一體化算力網。

顯然,算力已成為經濟高質量發展的重要動能,從過去“東數西算”的啟動到現在“全國一體化算力網”的構建,國家更加注重新算力基礎設施建設的質量。

“未來的算力,可能像現在的電力一樣,能夠實現全國的統一調配管理,即取即用。”李新宇對證券時報記者表示。

價格將趨於合理

值得一提的是,激增的算力需求以及美國高算力芯片供應受限等,已導致市場算力產品供應不足,價格暴漲,這在一定程度上增加了企業的算力成本。

vivo副總裁、AI全球研究院院長周周曾在接受證券時報記者採訪時坦言,“手機調用雲端大模型,目前大概一次1.5分,這個成本非常高,我們3億用戶每天用十次,一年下來要一百億元左右的支出。”

國盛證券《ChatGPT需要多少算力》報告顯示,大模型前期訓練成本很高,一次超過百萬美元,這個費用不僅涵蓋了模型的架構、算法和訓練數據的選擇,還包括了模型訓練所

需要的大量計算資源和時間成本。而隨着模型上線部署後使用量的增加,推理成本可能遠超訓練成本,尤其隨着大語言模型不斷升級迭代,參數量突破萬億、模型能力越來越強、用戶使用量激增,推理成本也會呈何式增長。

對此,劉宗長對證券時報記者表示,造成算力成本高的原因主要有兩方面,一是因為需求大幅增長而產能不夠,二是由於地緣政治、美國芯片禁令影響。

“目前公司正配合客戶在布局新的產能,預計明年在供應端會有一個明顯的提升,同時隨着供應的恢復,產品價格也會有所下降,進而會帶動更多新的需求。”劉宗長說,“當供應與需求開始平衡後,算力價格會逐漸趨於合理,回歸理性。”

李新宇也認為,隨着時間的推移,技術的進步、生態的成熟,未來算力成本會逐步降低。

算力成本較高另一原因,是我國算力產業仍與國際頭部企業存在一定差距,未來若能把握產業生態的發展規律,供給格局也將發生變化。

中國科學院院士王懷民對證券時報記者表示,我國算力產業發展要抓住機遇快速迭代升級,不僅要從芯片、軟件方面發力,更要注重生態的培育和打造,而中國在生態營造上具備獨特的優勢,中國具有龐大的消費市場、工程師群體等。