

行業概覽

除另有說明者外，本文件中本節及其他部分所呈列的資料及統計數據乃摘錄自不同的政府官方刊物及其他刊物，以及受我們就[編纂]委託的獨立市場研究及諮詢公司弗若斯特沙利文所編製的獨立行業報告。來自官方政府來源的資料未經我們、獨家保薦人、[編纂]、[編纂]、[編纂]、[編纂]、[編纂]、任何[編纂]、我們及彼等各自的任何董事及顧問，或參與[編纂]的任何其他人士或各方獨立核實，我們概不就其準確性發表任何聲明。

中國人工智能芯片市場

概要

近年來，人工智能(AI)已從實驗性技術轉變為主流應用，這得益於深度學習技術的突破、海量數據集的可用性以及計算能力的顯著進步。這項演進催化了AI在各行各業的廣泛應用——從自動駕駛汽車及金融服務到醫療保健及工業自動化——創造了對專用計算硬件前所未有的需求。大語言模型及電腦視覺應用的興起進一步加速這項轉變，將計算需求推向新高。

爆發式增長的需求催生了AI芯片的發展，AI芯片是專為AI應用設計的集成電路。該等芯片已成為現代算力基礎設施的關鍵組成部分，旨在高效處理AI算法所需的大規模並行計算任務，包括矩陣計算、模型訓練及推理過程。其架構代表著與傳統計算範式的根本性轉變，針對AI工作負載的獨特特性(例如並行計算、內存帶寬和能效)進行優化。

AI芯片類型

AI應用的密集計算需求推動處理器架構的根本性演變。雖然早期AI工作負載依賴傳統CPU，但其架構不足以滿足現代AI算法的並行計算需求。如今，兩種主要架構主導AI芯片市場：通用圖形處理單元(通用GPU)及專用集成電路(ASIC)，除通用GPU及ASIC外，AI芯片的其他輔助架構包括現場可編程閘陣列(FPGA)：

- 通用GPU乃為通用計算設計的圖形處理單元，意味著其能夠執行廣泛的計算任務，而不限於特定應用，其已成為主流的AI芯片架構，利用其並行計算能力及多功能性處理從模型訓練到推理任務的各種工作負載。其成熟的軟件系統、大量的開發工具及廣泛採用的編程框架已將通用GPU確立為AI算力基礎設施的基石。

行業概覽

- 採用算法嵌入式硬件架構的ASIC，雖然為特定任務提供卓越的性能及能效，但對不斷發展的AI算法的適應性有限，這可能會限制其更廣泛的應用。在ASIC中有多種架構，包括NPU(神經網絡處理單元)、TPU(張量處理單元)、DPU(深度學習處理單元)及IPU(基礎設施處理單元)等。
- FPGA指在製造後可重新配置其硬件功能的集成電路，在可編程性與硬件優化性能之間提供獨特的平衡。FPGA特別適用於要求低延遲的任務，例如實時推理、網絡加速及信號處理，或適用於最終算法或標準尚未固化的場景。FPGA主要部署在其靈活性及性能的獨特組合可提供決定性價值的特定應用中。

市場規模

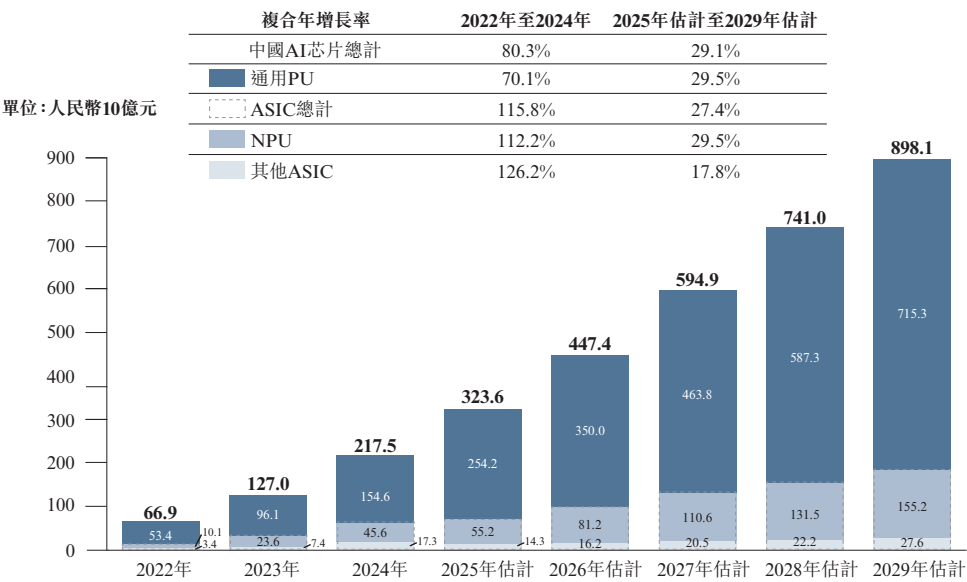
中國AI芯片市場經歷了顯著增長，這主要由對計算能力的爆炸性需求所驅動，特別是歸因於AI應用及大語言模型的普及，該等模型在雲端計算及邊緣側計算場景中均需要大量的計算資源。支持性政府政策促進技術研究及行業發展，同時國內半導體供應鏈的成熟及AI應用的成功商業化為行業建立了可持續的收入流，進一步催化這一增長。

近年來，中國已實施一系列支持性政府政策以促進其通用GPU行業的發展。例如，於2024年1月發佈的《工業和信息化部等七部門關於推動未來產業創新發展的實施意見》推動國產通用GPU在AI算力場景的應用，而於2023年8月發佈的《新產業標準化領航工程實施方案(2023–2035年)》則聚焦於建立獨立AI芯片標準及生態系統發展。於2023年8月發佈的《電子信息製造業2023–2024年穩增長行動方案》強調透過提供供應鏈強化及稅收優惠待遇支持，提升國產通用GPU企業的產能。該等政策透過構建支持技術進步、標準化及產能提升之綜合框架，共同推動行業增長。

行業概覽

2024年，中國AI芯片收入約為人民幣2,175億元，2022年至2024年的複合年增長率為80.3%，預計2029年將達到人民幣8,981億元，2025年至2029年的複合年增長率為29.1%。

中國AI芯片市場規模(按收入計)



資料來源： 相關上市公司年報、弗若斯特沙利文

附註： 鑑於FPGA目前僅佔整體AI芯片市場相對較小的比例（2024年<3%），為清晰與聚焦起見，本圖表並未將FPGA顯示為單獨區塊。

中國通用GPU市場概覽

概覽

通用GPU已成為AI應用的基礎算力核心組件，利用其龐大的並行計算能力、優化的內存架構以及在各種AI計算工作中廣泛共識的可拓展性。該等處理器最初是為圖形渲染而設計，現已發展成為AI計算的首選，原因是其能夠高效處理AI算法所特有的複雜矩陣計算及並行計算。

行業概覽

通用GPU行業由英偉達CUDA(統一計算設備架構)平台(一個全球主流通用GPU編程生態系統及平台)主導，其已成為最廣泛採用的加速計算平台。該平台如通用GPU的「操作系統」——其提供軟件工具，使開發人員能夠高效利用通用GPU計算能力。自2006年推出以來，全球開發人員一直依賴該系統編程及訓練AI模型，從而創造龐大的技術生態系統。即使出現新的通用GPU產品，兼容該平台對於應用廣泛推廣仍然至關重要。如今，超過80%的AI場景依賴該平台進行研發。隨著AI應用的擴展，對於通用GPU公司來說，與全球主流通用GPU編程生態系統及平台保持一致仍然有利於吸引開發人員社群，並對於保證硬件的即插即用至關重要。

於中國，蓬勃發展的AI市場直接推動通用GPU行業前所未有的增長，各行各業及應用領域的需求激增。市場的擴張主要受企業及消費者應用迅速採用AI技術驅動。例如，AI技術正被應用於金融服務領域，透過數據驅動智能轉變服務模式及強化風險控制系統；用於醫療保健領域，以實現智能疾病診斷、治療計劃及醫療資源分配；用於零售領域，以建立從需求預測到個人化服務的全棧智能系統；用於教育領域，以促進教育需求與靈活資源分配的精準匹配。中國通用GPU公司已通過開發出具有較強的計算核心、支持高精度計算及優化內存帶寬的芯片，從而滿足國內對高效AI計算基礎設施不斷增長的需求。

通用GPU芯片分類

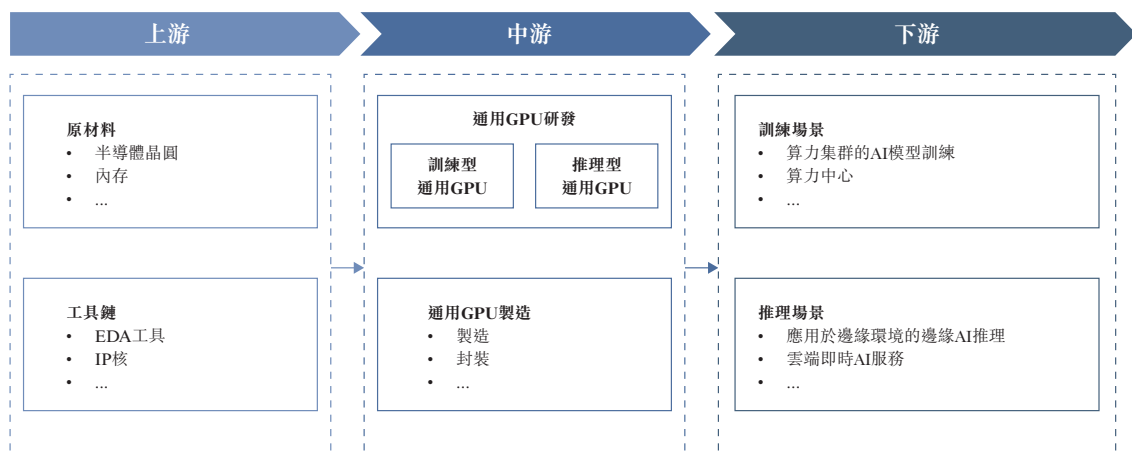
通用GPU市場已自然分為兩種類別，以應對AI計算生命週期的不同階段：訓練及推理。這種分類反映AI應用中不同的計算要求及部署場景。訓練通用GPU專為模型開發的密集計算需求而設計，而推理通用GPU則針對高效模型部署及執行進行優化。這種市場區分使通用GPU公司能夠開發專用產品，有效滿足特定的性能要求及用途，詳情如下：

特點	訓練	推理
主要功能	模型訓練，從大型數據集中學習	實時預測及模型部署
主要要求	— 大容量內存帶寬 — 並行計算能力	— 低延遲 — 能效
典型應用	— 大語言模型訓練 — 複雜模型開發	— 雲推理服務 — 邊緣計算

行業概覽

供應鏈

通用GPU憑藉其並行計算及高能效，已成為AI的基礎算力底座。為滿足AI爆發式增長的計算需求，通用GPU行業已形成涵蓋上游晶圓生產與設備、中游通用GPU設計與製造，以及下游AI專用場景部署（從模型訓練到邊緣設備推理）的產業鏈，實現AI工作鏈中的高效並行計算。



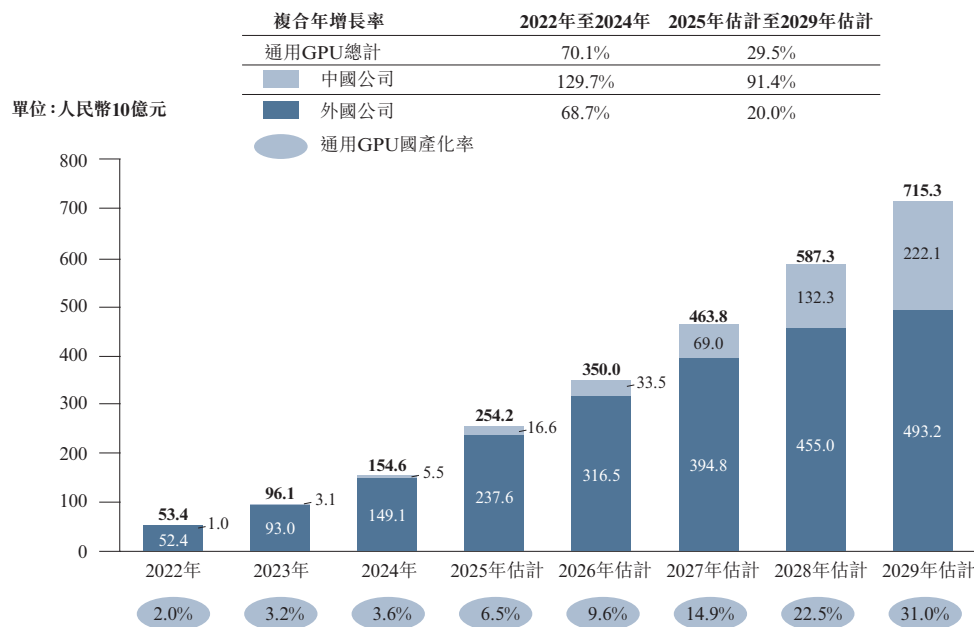
資料來源：弗若斯特沙利文報告

市場規模

受AI應用（特別是大語言模型）的爆炸性增長、雲計算基礎設施的迅速擴張以及國內芯片製造商的顯著進步所驅動，中國通用GPU市場的收入於2024年達到人民幣1,546億元，2022年至2024年的複合年增長率為70.1%。市場預計將保持強勁增長勢頭，到2029年收入將達到人民幣7,153億元，2025年至2029年的複合年增長率為29.5%。同時，國內通用GPU公司的收入增長顯著快於外國競爭對手。國產化率指國內通用GPU公司的收入佔中國通用GPU市場收入的比例。中國通用GPU市場的國產化率持續上升。從2022年到2024年，國產化率由2.0%增加至3.6%，預計到2029年將達31.0%。

行業概覽

中國通用GPU市場規模(按收入計)



資料來源：弗若斯特沙利文報告

中國通用GPU市場國產化率持續提升，乃受技術進步、市場需求強勁及政策大力支持所驅動。具體而言，製程節點與架構的不斷創新，加上生態兼容性的改善，使國產通用GPU企業得以縮小與全球領先者的性能差距。性價比的提升亦加快其市場接受程度。此外，AI應用的快速增長大幅擴張中國通用GPU市場。更多中國企業基於技術成熟及全球半導體貿易環境變化，正探索並採用國產通用GPU作為其採購策略的一環。此市場驅動之轉型促使需求增長，並形成更多元化及本土化之供應鏈格局。利好政策環境亦推動國產化率提高。國家戰略積極促進通用GPU自給自足，例如中國提出2025年芯片自給率達70%的目標。於2024年1月，工業和信息化部等七部門聯合發佈《關於推動未來產業創新發展的實施意見》，明確要求突破GPU技術以滿足大模型訓練及推理需求。地方政府透過補貼及獎勵措施，進一步加大對區域內通用GPU企業的支持力度。

行業概覽

市場驅動因素及未來趨勢

- **AI模型的複雜性提高及廣泛部署：**AI技術的快速發展，特別是大型基礎模型和特定行業垂類模型（例如醫療診斷、金融科技）的出現，創造了前所未有的計算需求。該等模型日益複雜且參數不斷擴大，需要龐大的算力資源進行訓練，而其於各行業的廣泛部署則推動了不斷增長的推理需求。這種演變正在改變市場動態，推理型通用GPU預計比訓練型增長更快。
- **多樣化的計算要求驅動通用GPU的廣泛應用：**AI算力需求負載的多樣性 — 從訓練大型基礎模型到運行專業垂類模型 — 需要靈活、性能優化的計算解決方案。通用GPU因其獨特的並行計算能力、架構靈活性及成熟的軟件系統組合而成為該等多樣化AI算力需求的首選解決方案。其能夠高效處理大型基礎模型的訓練以及各種垂直應用的推理，再加上持續的性能改進及廣泛的開發工具，已將通用GPU確立為AI工作負載的主流硬件加速卡。
- **集中式計算基礎設施發展：**AI模型參數的指數級增長使得傳統的通用GPU單卡或小規模計算集群不足以滿足高效的訓練和推理需求。中國的AI基礎設施正迅速轉向超大規模通用GPU算力集群，以支持萬億參數模型訓練，推動產品針對大規模並行計算和超高帶寬進行優化，並提高系統級穩定性。
- **邊緣計算及推理需求的加速：**隨著AI技術在各行各業中成熟及規模化，邊緣計算預計將經歷爆炸性增長，這得益於對國產化處理能力的需求。這種激增直接加速了對推理通用GPU的需求，推理通用GPU能夠實現實時決策、自主任務執行和對複雜環境的動態適應等關鍵功能。因此，推理通用GPU預計將比訓練通用GPU增長更快，到2029年將佔中國通用GPU市場的67%。這一趨勢為國內通用GPU公司帶來了巨大的機遇，因為他們正在為加速國產化替代努力。
- **國內通用GPU公司的技術進步：**通過架構設計和製造工藝的持續突破，國內通用GPU公司正在迅速提升其技術能力。這一進步，在國內市場大規模部署的促進下，技術進步使國內公司能夠顯著縮小與全球領先者的差距。在政策支持和不斷增強的市場競爭力的推動下，該等技術改進正驅動中國通用GPU市場的強勁增長，同時加速國產替代。

行業概覽

- **市場本地化程度不斷提高：**隨著國內通用GPU的技術持續進步，國內通用GPU公司正獲得更強的市場競爭力並穩步擴大其市場份額。2024年，通用GPU的國產化率超過3.6%。隨著技術成熟度和市場認可度的提高，國內通用GPU公司預計將進一步提升其市場份額。預計到2029年，通用GPU的國產化率將超過50%，從而在中國通用GPU市場形成更加多元化的競爭格局，並增強中國在全球通用GPU市場的影響力。

進入壁壘

- **技術複雜性：**作為性能優化的計算基礎設施的核心組件，通用GPU開發需要在架構設計、製造工藝整合及算法優化方面都具備核心技術能力。成功需要掌握多個技術領域，包括芯片設計、軟件棧開發及AI計算優化。這種多維度的技術複雜性，加上所需多年的研發投入，使得新進入者難以達到具競爭力的產品性能。
- **資本密集型要求：**通用GPU市場在整個產品生命週期中需要持續投入大量資本。公司須在EDA許可證等基本設計工具、先進的研發基礎設施以及各產品迭代的高昂流片成本上維持巨額支出。開發及生產方面持續投資的需求，加上為保持競爭力所需的頻繁產品迭代，構成巨大的資本壁壘，能夠承受的新進入者僅少數。
- **製造規模要求：**通用GPU市場要求建立先進的製造能力，或與頂尖晶圓代工廠建立穩固合作關係以獲取前沿工藝。尋求整合製造能力的新進入者需面對巨額工廠投資及技術專業知識要求；而採用無晶圓廠模式的公司則須在晶圓代工廠認證、產能分配及供應鏈管理等方面突破多重壁壘。該等要求，加之最低產能承諾及複雜的技術整合需求，為試圖進入市場的潛在競爭者構築重大壁壘。
- **研發人才與能力：**通用GPU市場以技術快速迭代為特徵，需要卓越的研究能力。在芯片架構、軟件開發及AI計算方面擁有專業知識的研發人員稀缺，構成重大壁壘。頭部企業憑藉穩定的協作團隊、長期積累的技術知識以及高效的開發流程，使得新進入者難以構建並維持具有競爭力的研發能力。

行業概覽

- **先發優勢與市場信任：**作為AI計算基礎設施的重要組件，通用GPU客戶對產品可靠性及性能驗證提出嚴格要求。先行企業受益於已建立的客戶關係、經證實的部署往績記錄以及累積的行業專業知識。藉此會創造一個自我強化的循環，成功的部署會深化客戶信任及忠誠度，形成新進入者難以突破的壁壘。

中國通用GPU市場競爭格局

目前外國公司的通用GPU在中國通用GPU市場就收入而言仍佔據主要，而中國公司的預期市場份額將不斷增加。2024年，按收入計，包括本公司在內的三家中國公司已躋身中國通用GPU市場五大參與者之列。2024年中國通用GPU市場參與者按收入的排名如下表所示。

排名 ⁽⁶⁾⁽⁷⁾	公司	中國公司	2024年	2024年
			中國通用GPU 市場份額	國內通用GPU 市場份額 ⁽¹⁾
1	公司A ⁽²⁾	×	91.9%	—
2	公司B ⁽³⁾	×	4.5%	—
3	公司C ⁽⁴⁾	√	0.8%	23.3%
4	公司D ⁽⁵⁾	√	0.5%	13.5%
5	本公司	√	0.3%	9.8%

資料來源：弗若斯特沙利文報告

附註：

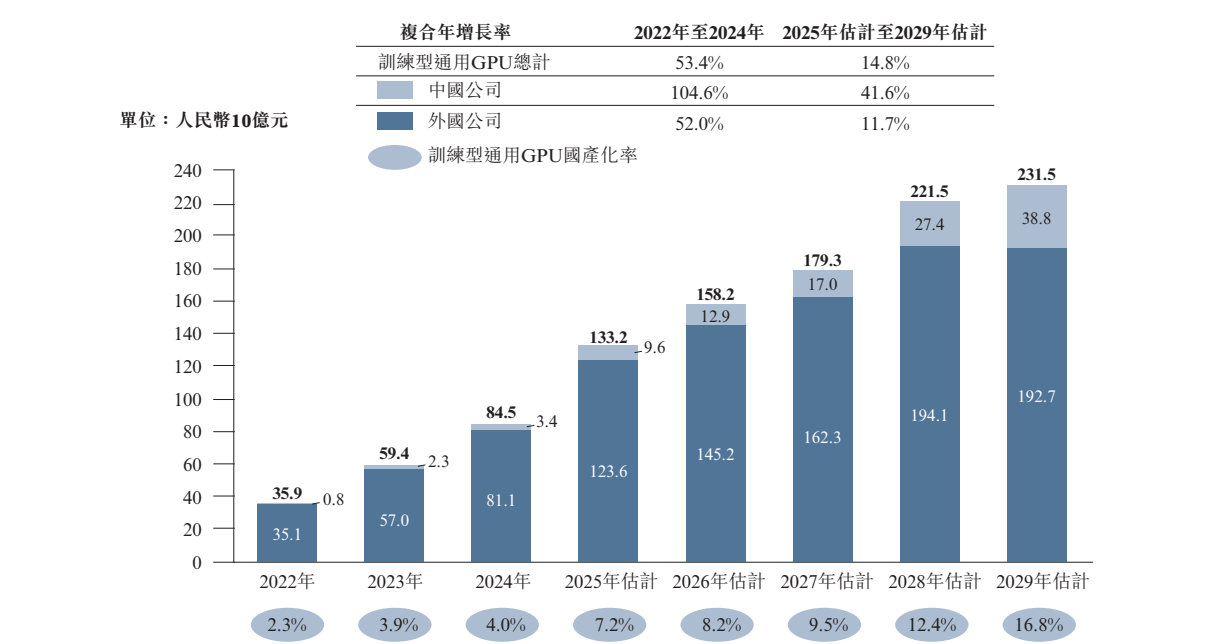
- (1) 國內通用GPU市場份額指中國公司於中國通用GPU市場中通用GPU的收入中所持有的市場份額。
- (2) 公司A為一家專注於GPU及AI計算平台開發的科技公司，成立於1993年，總部位於美國，並於納斯達克上市。
- (3) 公司B為一家專注於微處理器、GPU及AI算力解決方案設計的半導體公司，成立於1969年，總部位於美國，並於納斯達克上市。
- (4) 公司C為一家專注於高性能處理器開發及銷售的科技公司，成立於2014年，總部位於中國，並於上海證券交易所市場上市。
- (5) 公司D為一家專注於通用GPU及AI算力解決方案開發的GPU設計公司，成立於2020年，總部位於中國，並擬於上海證券交易所市場上市。
- (6) 公司A至D的收入數據基於單位銷售額及其他可得公開資料估算，並經與業界人士及專家盡力核對，因此可能與上述公司的經審核數據存在差異。
- (7) 可得公開資料包括上市公司財務報告及公開行業報告等。

行業概覽

中國訓練型通用GPU市場規模

在2022年至2024年的AI發展初期階段，AI模型訓練的龐大需求推動訓練型通用GPU主導中國通用GPU市場。2024年，中國訓練型通用GPU市場收入達人民幣845億元，2022年至2024年的複合年增長率達53.4%，表現突出。展望未來，預計2025年至2029年，中國訓練型通用GPU收入將以14.8%的複合年增長率增長，到2029年達到人民幣2,315億元。同時，訓練型通用GPU的國產化率預計將由2024年的4.0%攀升至2029年的16.8%。

中國訓練型通用GPU市場規模(按收入計)



資料來源： 弗若斯特沙利文報告

行業概覽

中國訓練型通用GPU市場競爭格局

目前外國公司主導中國訓練型通用GPU市場，但預計未來數年中國公司的市場份額將大幅提升。2024年中國訓練型通用GPU市場五大參與者中，四名參與者為中國公司。2024年中國訓練型通用GPU市場參與者按收入的排名如下表所示。

排名 ⁽³⁾⁽⁴⁾	公司	中國公司	2024年 中國 通用GPU 市場份額	2024年 國內 通用GPU 市場份額 ⁽¹⁾
1	公司A	×	96.0%	—
2	公司C	√	1.1%	26.3%
3	公司D	√	0.7%	17.4%
4	本公司	√	0.4%	10.5%
5	公司E ⁽²⁾	√	0.3%	7.9%

資料來源：弗若斯特沙利文報告

附註：

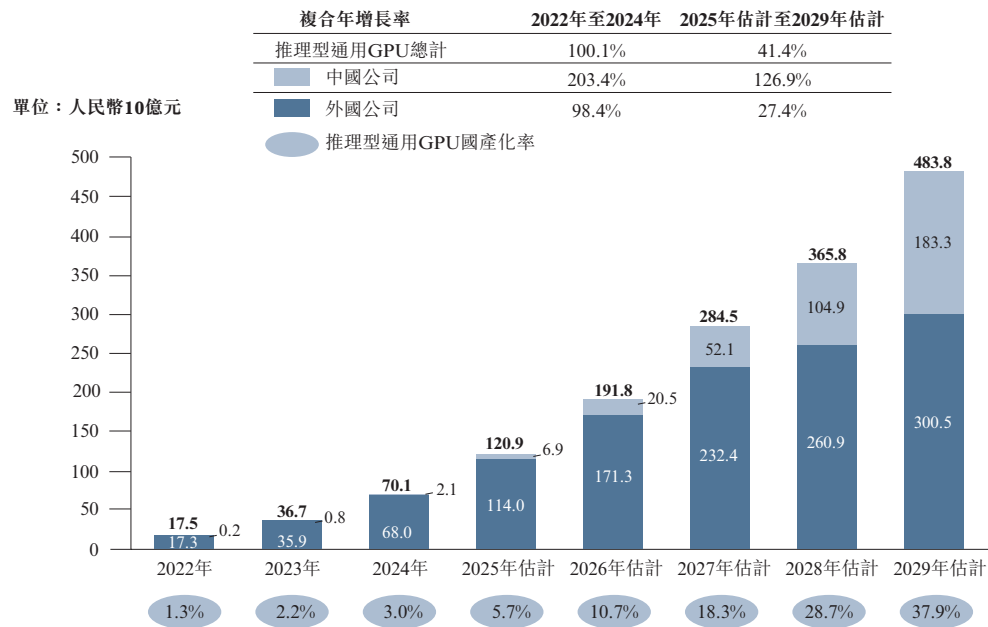
- (1) 國內訓練型通用GPU市場份額指中國公司於中國通用GPU市場內訓練型通用GPU的收入中所持有的市場份額。
- (2) 公司E為一家專注於通用GPU及AI算力解決方案設計的AI算力解決方案提供商，成立於2019年，總部位於中國。
- (3) 公司A至E的收入數據基於單位銷售額及其他可得公開資料估算，並經與業界人士及專家盡力核對，因此可能與上述公司的經審核數據存在差異。
- (4) 可得公開資料包括上市公司財務報告及公開行業報告等。

中國推理型通用GPU市場規模

2024年，中國推理型通用GPU市場收入為人民幣701億元，2022年至2024年的複合年增長率達100.1%，表現突出。展望未來，隨著AI應用在產業端更廣泛落地，預計推理型通用GPU的市場份額將不斷增加並主導市場，反映了產品的技術成熟度與商業化進展。預計2025年至2029年，中國推理型通用GPU收入將以41.4%的複合年增長率增長，到2029年達到人民幣4,838億元。同時，推理型通用GPU的國產化率預計將由2024年的3.0%攀升至2029年的37.9%。

行業概覽

中國推理型通用GPU市場規模(按收入計)



資料來源： 公開資料、專家訪談、弗若斯特沙利文

中國推理型通用GPU市場競爭格局

目前，儘管外國公司在中國推理型通用GPU市場佔據主導地位，但隨著推理型通用GPU市場快速增長及中國公司技術競爭力加強，中國公司有望取得顯著市場影響力。未來幾年，中國公司預計將大幅增加市場份額。2024年中國推理型通用GPU市場參與者按出貨量的排名如下表所示。

排名 ⁽²⁾⁽³⁾	公司	中國公司	2024年	2024年
			中國通用GPU市場份額	國內通用GPU市場份額 ⁽¹⁾
1	公司A	×	87.0%	—
2	公司B	×	10.0%	—
3	公司C	√	0.5%	18.5%
4	本公司	√	0.3%	8.6%
5	公司D	√	0.2%	7.0%

資料來源： 弗若斯特沙利文報告

行業概覽

附註：

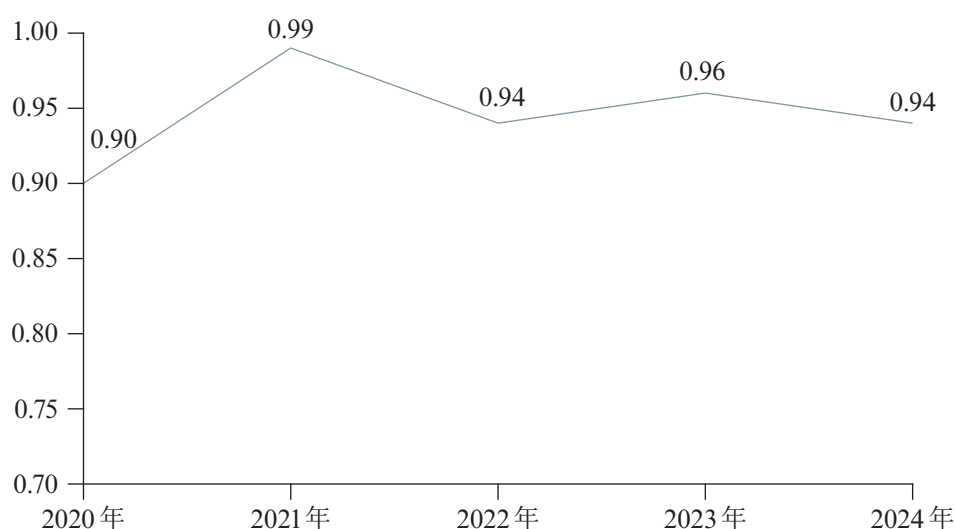
- (1) 國內推理型通用GPU市場份額指中國公司於中國通用GPU市場上推理型通用GPU的收入中所持有的市場份額。
- (2) 公司A至D的收入數據基於單位銷售額及其他可得公開資料估算，並經與業界人士及專家盡力核對，因此可能與上述公司的經審核數據存在差異。
- (3) 可得公開資料包括上市公司財務報告及公開行業報告等。

半導體硅片的歷史價格

半導體硅片為通用GPU的材料之一。於2020年，全球半導體芯片需求不斷增加，半導體硅片的價格由每平方英吋0.9美元提升至每平方英吋0.99美元。自2021年起，半導體硅片的價格隨著產能擴張呈現下降趨勢。考慮到全球產能充足，預測半導體硅片的價格長期將保持穩定。

半導體硅片的價格

單位：每平方英吋美元



資料來源：國際半導體產業協會(SEMI)、弗若斯特沙利文報告

資料來源

就[編纂]而言，我們委聘獨立市場研究顧問弗若斯特沙利文，對全球及中國的AI及通用GPU市場進行分析，並編製弗若斯特沙利文報告。弗若斯特沙利文是一家獨立的全球諮詢公司，於1961年在紐約成立，其服務包括(其中包括)行業諮詢、市場戰略諮詢及企業培訓。就提供的市場研究服務而言，我們已向弗若斯特沙利文支付人民幣600,000元的費用，我們認為該費用符合市場費率。

行業概覽

於編纂及編製弗若斯特沙利文報告時，弗若斯特沙利文進行了(i)初步研究，包括訪談行業參與者、競爭對手、下游客戶及公認的第三方行業協會；以及(ii)次要研究，包括審閱公司年報、相關官方機構的數據庫，以及弗若斯特沙利文在過去幾十年建立的獨家數據庫。弗若斯特沙利文報告中的市場預測基於預測期內的以下關鍵假設：(i)所討論的全球及中國市場的社會、經濟及政治狀況在預測期內將保持穩定；(ii)全球及中國市場的政策在預測期內將保持一致；以及(iii)全球及中國市場將由弗若斯特沙利文報告中所述的因素驅動。

除非另有說明，本節所載的所有數據及預測均來自弗若斯特沙利文報告。該委託報告由弗若斯特沙利文獨立編製，不受本公司或其他利益相關方的影響。我們的董事確認，據彼等作出合理查詢後所知，自弗若斯特沙利文報告日期以來，市場資訊並無重大不利變動，而該等變動可能規限、抵觸或影響本節所載資訊。