

業 務

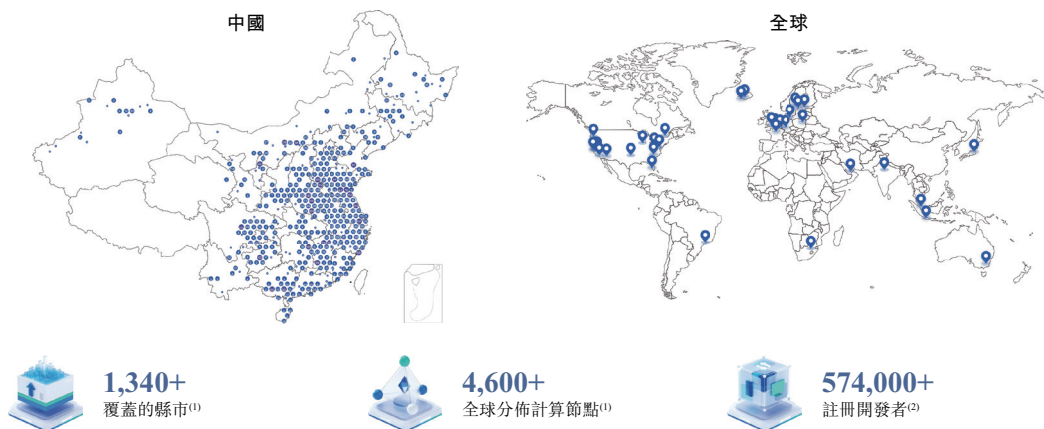
概覽

我們是中國領先的獨立分佈式雲計算服務商。根據灼識諮詢的資料，截至2025年12月31日，按計算節點數計，我們運營著中國最大的算力網絡，按2025年收入計，我們是中國最大的獨立邊緣雲計算服務商。按2025年日均token消耗量計，我們在中國獨立AI雲計算服務商中亦排名第一。

我們的業務始於邊緣雲計算服務領域，因為我們發現海量的算力需求與分散且尚未充分利用的算力供給之間存在重大錯配。我們搭建了算力調度平台，匯聚和管理異構資源的閒置算力，並以高效方式向客戶部署，從而加速內容傳輸，提升快速回應、低延遲計算體驗。2023年，隨著大模型的快速發展，帶來圖形處理單元（GPU）算力及AI推理需求激增，我們戰略性地開展AI雲計算服務。依託我們的全球算力調度網絡及綜合AI原生能力，我們將異構、跨區域的計算資源轉化為低成本、低延遲及高度可擴展的計算服務，服務全球AI原生應用程序、開發者及企業。我們的核心價值並非單純提供計算資源的存取權限，而在於提供可靠、高效及優化的模型服務，讓客戶能夠將其無縫整合至彼等的產品及工作流程中，而無需構建或管理底層基礎設施。我們亦正將我們的平台擴展至「智能體基礎設施」，創建將為AI智能體自主及持久執行任務提供動力的基礎層。隨著智能體AI的普及，我們預期雲計算基礎設施的主要調用者將由人類用戶轉向AI智能體，從而產生對計算資源的全新需求。

截至2025年12月31日，我們的算力網絡覆蓋全球1,340多個縣市，包含4,600多個計算節點。截至2026年4月30日，我們AI雲計算服務在全球擁有超過574,000名註冊開發者。根據灼識諮詢的資料，於2026年4月，我們的日均token消耗量達10,280億次，使我們成為中國最大的獨立AI雲計算服務商。

我們的算力網絡



附註：

- (1) 截至2025年12月31日。
- (2) 截至2026年4月30日。

業 務

我們的技術實力及服務能力贏得了業界的廣泛認可。我們為眾多國內外領先科技公司（根據灼識諮詢的資料，包括大部分的中國前十大互聯網公司）、雲計算公司以及「獨角獸」初創公司提供邊緣雲計算服務和AI雲計算服務。

於往績記錄期，我們的收入由2023年的人民幣358.4百萬元迅速增長至2024年的人民幣558.0百萬元，並進一步增長至2025年的人民幣770.3百萬元，2023年至2025年的複合年增長率為46.6%。特別是，自2025年起，我們的AI雲計算服務出現爆炸式增長，收入由2024年的人民幣10.4百萬元增加逾十倍至2025年的人民幣119.2百萬元。

我們的市場機遇

AI的快速迭代帶來了龐大的算力需求。儘管供應增加，但由於計算資源的異構性、地域分佈廣泛及時間錯配，全球算力的利用效率仍然較低，平均利用率在50%至70%。雖然獲取額外計算資源十分重要，但更大的挑戰在於將該等分散的算力轉化為穩定、商業級的服務。

憑藉靈活的資源配置及輕資產運營模式，分佈式雲計算平台可有效突破該等局限。尤其是，針對AI計算需求，分佈式雲平台提供對跨地區分散、多代計算資源的標準化接入，將其整合為一個統一、可管理的算力池。利用需求的地域及時間差異（例如，西半球的午夜通常為非繁忙時間，而同時東半球的白天則為繁忙時間），該平台動態地將未充分利用的計算資源調度至高需求的推理任務，從而持續地最大化整個網絡中資源的整體利用率。分佈式雲計算近年來呈現爆炸式增長。根據灼識諮詢的資料，按收入計，2025年全球邊緣雲計算服務市場為人民幣2,098億元；2025年中國邊緣雲計算服務市場達人民幣153億元，預計2030年將接近人民幣419億元；2025年全球AI雲計算服務市場達人民幣1,882億元，預計2030年將接近人民幣46,668億元，複合年增長率為90.1%。

AI的發展及大模型的應用正革新日常生活及工作場景。儘管AI開發者已投入大量算力訓練模型，使其具備更高的能力及複雜程度，但現時越來越多的資源用於AI「推理」，即利用訓練好的AI模型針對應用場景生成分析結果的過程。目前，AI推理在算力消耗方面已超越AI訓練。AI推理應用程序大規模滲透至消費者及企業的工作流程中，對高效、彈性及具成本效益的推理服務產生持續、高頻的需求。展望未來，隨著AI從被動生成答案過渡至全天候主動執行複雜的多步驟任務，智能體AI時代正在興起。隨著AI智能體的普及，雲計算基礎設施的主要調用者將由人類用戶轉向自主發起任務的AI智能體，從而產生對計算資源的全新需求，以支持多步驟、工具密集型的智能體工作負載。我們認為，長遠而言，算力需求增長的主要驅動力將為智能體AI任務，因為智能體AI將日益滲透至人們的生活及工作場景中，並產生大規模的算力需求。基於同樣原因，我們認為企業及個人用戶對高彈性及具成本效益的算力服務將存在巨大需求。

業 務

我們的產品及服務



誠如上圖所示，我們已構建一個跨越三個無縫整合層的雲計算平台。其基礎為我們的全球算力網絡，或基礎設施即服務 (IaaS) 層，該網絡將全球分散且異構的計算資源統一為高效、可擴展且標準化的計算底座。在此之上，MaaS (模型即服務) 層提供對即用型模型的存取，並透過專有推理優化技術加以提升，該等技術顯著提高其性能，超越了開箱即用的能力。隨著AI邁向智能體時代，我們正進一步擴展至智能體基礎設施層，為能夠自主執行高頻任務的AI智能體奠定基礎。該等層次共同組成一個具凝聚力的平台，旨在為下一代大規模AI創新提供動力。

具體而言，我們於邊緣雲計算服務及AI雲計算服務分部下提供以下產品：

- **邊緣雲計算服務。**我們的邊緣雲計算服務整合第三方提供商的分散異構計算資源，滿足客戶對邊緣計算及傳輸的需求。我們的邊緣雲計算服務並非將數據傳輸至集中式雲數據中心，而是將算力下沉到互聯網「邊緣」(即靠近數據生成及使用的位置)處理。此功能對於需要實時高速處理海量數據的客戶(如音視頻內容提供商)至關重要，乃由於低時延是良好用戶體驗的關鍵。具體而言，我們提供以下服務：
 - **邊緣節點服務。**利用我們廣泛分佈的邊緣計算節點網絡，邊緣節點服務將算力從中心節點遷移至更靠近用戶的邊緣節點，旨在支持邊緣存儲、音頻及視頻處理、視頻轉碼及分發等場景。具體而言，我們的產品包括邊緣容器及邊緣裸金屬。邊緣容器是部署在邊緣的虛擬化單元，允許應用程序在數據生成位置或附近高效運行，從而降低時延並提高實時應用程序的性能。邊緣裸金屬是指部署在邊緣的獨立物理服務器，為計算提供專用硬件資源。

業 務

- **邊緣內容分發網絡(邊緣 CDN)**。我們的邊緣CDN依託廣泛的邊緣節點網絡，並利用我們在智能調度、分佈式緩存及網絡傳輸技術方面的專業知識。其主要應用場景包括文件下載、視頻點播及直播。不同於傳統CDN，我們的邊緣CDN與本地網絡更深度集成，使服務器更靠近終端用戶。這種近距離部署提升了傳輸質量，為用戶帶來更流暢的視頻播放及更快速的內容分發體驗。
- **AI雲計算服務**。利用集成在我們算力網絡中的分佈式GPU計算資源，我們於2023年推出了AI雲計算服務。隨著AI模型的規模和複雜性不斷提高，我們全面的產品組合滿足了AI推理領域的廣泛需求。這包括提供大型企業所需的彈性GPU計算資源，以及為中小型企業(SME)和開發者量身定制的多樣化模型服務。我們的服務旨在降低全球客戶在採用新技術方面所面臨的成本及技術壁壘。具體而言，我們提供以下涵蓋從基礎到高階功能的產品：
 - **GPU雲服務**。我們的GPU雲服務為客戶提供強大的計算資源，無需購買物理GPU或管理GPU服務器。該架構具有高度彈性，可按需自動擴縮容，採用按需付費模式，客戶只需按實際使用量付費。我們的服務不僅整合了GPU資源，更提供專為生成式AI和機器學習工作負載設計的強大運營環境。我們GPU雲服務的主要特性包括優化的推理引擎、預置模板快速部署及需求驅動的自動擴縮容。這些特性使我們的GPU雲服務成為AI推理的全方位平台，為開發和擴展AI應用程序提供高效性、靈活性及可靠性。
 - **模型API**。我們的模型API服務是客戶訪問預構建、開箱即用模型（主要是大語言模型(LLM)及多模態模型）的網關，專為自身缺乏構建、訓練或管理高度複雜AI模型能力的客戶設計。我們提供兩類服務：其一，我們的模型API直接接入領先開源模型，客戶可直接便捷地調用這些開源模型；其二，對於傾向使用自主開發模型的客戶，我們可將有關模型託管在我們的基礎設施上。這使我們的客戶能夠保留對其數據及模型行為的控制權，同時享受我們基礎設施的可擴展性、高效性與可靠性。

我們的模型API服務的一個顯著特徵是，我們並非單純提供通往底層模型的傳遞網關。我們的平台整合了一套專有的推理優化技術，該等技術實質上提升了託管模型的性能，超越了其開箱即用的能力。這使客戶能夠受惠於較標準模型部署明顯更快、更高效及更具成本效益的模型服務。例如，根據我們於2026年的評估，透過我們的模型API服務交付的一款知名大模型的性能，大幅超越使用SGLang（一個基準模型推理引擎）部署的同一模型，將首字節延遲降低95%，並將總吞吐量效率提高300%。

業 務

- **智能體基礎設施。**AI智能體是透過推理、規劃及採取行動來自主完成任務的AI系統，無需持續的人類監督。智能體基礎設施是一套雲端基礎設施服務，旨在為AI智能體提供可自主運行的雲計算環境，反映出日後AI智能體（而非人類）將成為雲服務主要使用者的趨勢。此項全面技術系統涵蓋AI編程、沙箱執行環境、廣泛的工具與技能，以及短期與長期記憶系統。

競爭優勢

我們相信，以下競爭優勢對我們的成功發揮關鍵作用，使我們從競爭對手中脫穎而出。

中國領先的獨立分佈式雲計算服務商

我們是分佈式雲計算服務行業的領導者、開拓者及創新者：

- **領導者。**我們是中國領先的獨立分佈式雲計算服務商。根據灼識諮詢的資料，截至2025年12月31日，按計算節點數計，我們運營著中國最大的算力網絡，按2025年收入計，我們是中國最大的獨立邊緣雲計算服務商。按2025年日均token消耗量計，我們在中國獨立AI雲計算服務商中亦排名第一。
- **開拓者。**我們的創始團隊由一群擁有逾20年經驗的分佈式系統開拓者組成。由於發現海量需求與分散且常未充分利用的供給之間存在重大錯配，我們是業內最先搭建算力調度平台的企業之一，匯聚和管理異構資源的閒置算力，並以高效方式向客戶部署，從而加速內容分發及提升低時延計算體驗。
- **創新者。**為應對生成式AI時代AI大模型的發展，我們對平台及服務進行創新。這一發展推動對GPU算力及AI推理能力的需求。我們通過將AI能力集成至我們的算力網絡中，推出AI雲計算服務。我們的AI雲計算服務為客戶提供生成式AI時代所必需的AI算力平台及一系列AI推理工具，例如高性能推理引擎、分佈式雲平台及模型API。我們是中國分佈式雲計算市場中最早提供AI雲計算服務的參與者之一。

生成式AI時代的到來，推動巨大的算力需求，企業用戶及個人開發者對低成本、高度靈活的AI推理算力的需求穩步增長。憑藉我們在分佈式雲計算行業的經驗以及在AI雲計算領域的戰略定位，我們認為我們具備先發優勢，能夠把握AI帶來的廣闊市場機遇。

高效且高度可擴展的業務模式推動自然驅動式增長

我們的算力網絡是我們運營的核心，連接我們業務生態系統中的各利益相關方，包括算力提供商、用戶及AI開發者。我們受益於這種高效且高度可擴展的平台業務模式，推動自然驅動式增長：

- **自然驅動式增長。**我們的平台匯聚全球計算資源，賦能全球用戶及AI開發者，有效解決算力市場錯配的問題，推動平台於往績記錄期快速增長。我們算力網絡中的計算節點數量由截至2023年12月31日的超過2,800個增至截至2024年12月31日的超過4,000個，並進一步增至截至2025年12月

業 務

31日的超過4,600個。隨著我們的業務增長，我們的收入由2023年的人民幣358.4百萬元快速增長至2024年的人民幣558.0百萬元，並進一步增長至2025年的人民幣770.3百萬元，2023年至2025年的複合年增長率為46.6%。

- **高效運營。**我們採用輕資產業務模式，將資本支出降至最低，這使我們能夠靈活地將更多資源投入研發活動，增強創新能力及快速響應市場變化，同時持續提升盈利能力。
- **高度可擴展性。**我們的分佈式雲計算服務平台具備快速、無縫的可擴展性，可實現跨區域高效複製。憑藉互聯算力網絡，我們的業務從部分地區快速擴展至中國幾乎所有地區。這種可擴展性可優化資源利用，同時滿足全球不斷增長的算力需求。我們快速複製及擴張的能力有助於建立一個彈性生態系統，滿足AI開發者及終端用戶不斷增長的需求。

堅實的技術基礎及強大的研發能力

在雲計算服務不斷發展的背景下，我們精心開發核心技術，旨在增強靈活性、可擴展性及易用性。秉承對兼容性及獨立性的承諾，我們的分佈式雲計算平台優勢突顯，能夠提供定制化尖端解決方案，滿足生成式AI時代自適應、多樣化需求。我們擁有以下核心技術：

- **自主研发的分佈式計算架構。**我們運營的核心是一個強大的分佈式算力調度平台，可無縫管理龐大的計算節點網絡。我們的系統將多樣化的計算資源整合為統一資源池。通過智能工作負載調度，我們的架構可降低底層硬件異構性，實現多樣化計算資源的無縫集成，並為上層應用提供標準化接入點。此外，基於輕量級雲原生架構，我們的系統沿著兩個互補的維度高效地組織碎片化的計算資源：人工智能計算層，處理人工智能推理及高性能計算工作負載；邊緣計算層，通過分佈式邊緣節點網絡驅動對延遲敏感的服務。此設計可讓工作負載根據性能要求、延遲時間限制及成本目標智能地路由至最佳資源層，同時較傳統解決方案可大幅減少系統開支。此外，通過機器學習技術，我們的自動故障診斷系統可預測潛在節點故障並實時動態調整任務冗餘。
- **算力網絡及調度能力。**我們已建成分佈式算力網絡，而高密度計算節點部署使本地分佈實現毫秒級時延（短至10毫秒），為實時音視頻內容分發、AI推理等場景提供基礎設施支撐。我們的全球調度網絡使我們能夠持續、全天候地實現計算資源的高利用率。通過利用跨地區時區差異和我們多樣化客戶群的交錯需求概況，我們將全球閒置的計算能力動態分配至任何特定時刻的需求高點，確保我們的GPU資源得到有效部署而不是在非高峰時

業 務

段閒置。因此，根據灼識諮詢，我們的平均GPU利用率於2025年一直高於75%，大幅超越行業平均水平40%至50%。這種智能的「削峰填谷」能力大幅提高我們網絡的整體資源利用率，直接降低向客戶交付每token的實際成本。

- **大模型的分佈式推理加速。**我們的平台包含一套專有的推理優化技術，可大幅提升託管模型的性能，超越其開箱即用的功能。這使我們的客戶能夠受益於明顯更快、更高效且更具成本效益的模型服務。我們的優化能力涵蓋算子優化、推理引擎優化、存儲系統優化、為前沿硬件量身定制的模型優化等多個維度。通過該等多層次的優化，我們能夠將計算資源利用率推至接近理論極限，從而大幅降低成本。例如，於2026年，通過我們的模型API服務交付的知名大模型的輸入代幣定價較主要國際雲服務提供商的平均收費低40%以上。
- **異構計算資源的整合。**與依賴集中自有、同質計算資源的傳統雲計算提供商的資本密集型模式不同，我們採用專有技術，使我們能夠將異構計算資源（例如，不同品牌、世代及位置的GPU）編排為一個統一的計算資源池，並為客戶提供標準化的訪問權限。與傳統提供商所依賴的高端、同質計算資源相比，這種計算資源組合可提供顯著更高的成本效益，同時仍能交付足夠的服務質量及可靠性。透過大規模整合該等多樣化且具成本效益的資源，我們大幅降低了基準單位計算成本，從而直接支持盈利能力的提升。

我們的研究已為分佈式雲計算的關鍵方面帶來重大的技術進步，如異構邊緣環境中的工作負載預測及故障感知，以及智能資源調度及情境感知任務調度。通過與領先的研究機構合作，我們已在頂級期刊或學術會議（如IEEE INFOCOM（國際電氣與電子工程師協會－國際計算機通信會議），一個專注與計算機通信有關的重要主題及問題的領先會議）上發表或提交18篇高影響力論文。

充滿活力並促進增長的社區

基於我們的算力調度平台及算力網絡，我們創建了一個有效連接閒置計算資源與計算需求的社區。圍繞我們平台的是一個充滿活力的社區，由算力用戶、AI開發者、算力供應商及其他業務合作夥伴（如研究機構和行業組織）組成。

對於我們的邊緣雲計算服務，我們將供應商的閒置計算資源與客戶的計算需求有效連接。對於算力供應商，我們的平台解決了行業內長期存在的挑戰：算力資源的高閒置率。根據灼識諮詢的資料，全球IDC的平均利用率僅為50%至70%左右，且對於其他類型的算力供應商而言可能更低，導致大量算力資源未充分利用。我們的服務對於釋放這一潛力至關重要，使供應商能夠將其閒置資源變現，並並以最少的運營調整創造新的收入來源。我們為客戶提供滿足其多樣化需求的高性價比、低時延及高彈性的雲計算服務及解決方案。通過使用我們的平台，客戶可節省構建維護自有數據中心所需的大量成本及避免有關複雜性。此外，我們的彈性服務可根據需求波動實時調整資源，客戶僅需按實際算力用量付費，優化資源利用率。

業 務

在AI雲計算服務方面，我們與擁有廣泛開發者網絡的全球開源社區（如GitHub及OpenRouter）合作。我們相信，開源社區是推動AI行業最前沿進步的主要引擎，而與該等社區保持深入、積極的互動對於保持AI技術的前沿至關重要。通過將我們的產品整合至合作夥伴生態系統及推行聯合市場推廣計劃，我們預期可加速獲取新客戶、擴大行業覆蓋範圍並提高每名客戶收入。我們於2026年1月在OpenRouter的第三方開源大模型提供商中排名第一，以及於2025年12月在Hugging Face的推理服務月請求量排名第一。我們亦實施開發者友好策略，以相對較低的獲客成本吸引新客戶。我們積極參與開源社區活動，提供開源整合，並舉辦線上及線下技術研討會及開發者論壇。於2025年，我們的AI雲計算服務擁有超過180,000名活躍用戶。

此外，我們已與多家學術機構及行業組織進行合作研究或戰略聯盟，以增強我們社區的多樣性及活力。例如，我們與多家領先的研究機構及企業組成了未來算力網絡協同調度聯盟。該聯盟乃為中國行業參與者搭建一個協作發展平台而設，以建立一個具有高算力、高傳輸容量及高成本效益的廣泛網絡。

對環境可持續發展及社會責任的承諾

AI計算屬高度能源密集型。隨著AI工作負載的複雜性和規模不斷增長，相關的能源消耗亦相應增加，從而導致運營成本和環境影響增加。自成立以來，我們一直將可持續發展作為我們業務模式及企業精神的核心原則。我們的業務本質上是以節約能源、減少碳足跡及履行社會責任為基礎。

傳統的IDC往往面臨能源利用效率低下的困境。許多IDC為了追求規模經濟，在單一計算節點下維護大量服務器，導致服務器冷卻消耗大量能源。運行效率低下的IDC可能將其總能耗的30%以上用於服務器冷卻，導致PUE值較高。相比之下，我們的分佈式平台可以連接較小計算節點與戰略性分佈在不同環境的節點。這種方法有助減少對PUE值較高的計算資源的依賴，最終提升整體平台能源效能。

此外，我們的分佈式雲計算服務通過優化算力調度促進環境可持續發展。例如，我們可將對時延不敏感的計算任務（那些不需要立即接近終端用戶的計算任務）分配給位於高緯度地區的計算節點（如我們在冰島及挪威的計算節點），利用當地自然較低的溫度以減少服務器冷卻所需的能耗。此外，我們亦可直接將任務分配至可再生能源來驅動算力，從而進一步提升我們的環境效率。

同時，我們的分佈式雲計算網絡可充分利用現有的、利用率不高的計算節點，因此減少額外建立海量集中式數據中心的需要。我們的技術亦有效延長計算設備的使用壽命並減少不斷替換或擴展集中式設施的需要。這減少了電子垃圾及與生產及處置計算設備有關的碳足跡。

我們還通過支持當地社區來體現我們對社會責任的承諾。位於低線城市的數據中心為當地經濟作出重大貢獻，成為技術發展及基礎設施改善的催化劑。通過最大限度地利用該等本地數據中心的算力，我們的網絡不僅提高了該地區的生產力，亦促進了社區發展。

業 務

環境可持續發展及社會責任是我們企業價值觀不可或缺的一部分。我們始終以卓越的環境管理為目標，並致力於履行我們的社會責任，努力成為雲計算行業的領導者，致力於創造更美好的未來。

管理團隊經驗豐富，擁有廣泛的行業背景

我們的管理團隊經驗豐富，擁有雄厚的行業背景及創業經驗。我們的核心管理團隊擁有超過20年的專業知識、強大的技術能力及商業化技能。

我們的首席執行官姚欣先生是一位連續創業者及投資人，在建立全球規模的技術平台及推動前沿技術投資方面擁有豐富經驗。作為PPTV的創始人，彼於2004年首創了全球首個點對點串流媒體協議，並將其發展成為中國首屈一指的互聯網電視平台。從PPTV成功退出後，彼成為藍馳創投（一家領先的技術型風險投資基金）的風險投資合夥人，彼專注於全球AI領域的早期投資。姚先生的全球視野推動了我們的商業及戰略擴張。

我們的首席技術官王聞宇先生是一位在技術上成就卓著的連續創業者。彼共同創立了PPTV並擔任首席技術官，開發了其分佈式視頻串流服務，並開創了多項點對點的音視頻流媒體的分發技術。在第二次創業期間，王先生作為聯合創始人共同創立並擔任極豆首席技術官，設計了軟硬件一體化的智能座艙及中國首個汽車智能操作系統。王先生發表了15篇關於邊緣計算、人工智能及大模型的論文及擁有18項專利。

首席科學家王曉飛教授是邊緣計算與AI系統領域的領軍人物。彼發表了200多篇高影響力的研究論文，引用次數超過11,000次。王教授亦獲得Fred W. Ellersick獎(Fred W. Ellersick Prize)，為IEEE通信學會為表彰優秀論文而頒發的著名獎項。其開創性著作《邊緣智能》是該領域最早且最具影響力的出版物之一。

技術委員會主席金海教授為華中科技大學教授，是分佈式及並行計算與雲計算系統領域的開拓者。金教授擔任服務計算技術與系統教育部重點實驗室主任，並擔任中國算力網（覆蓋全中國的數字基礎設施，實現國家大型算力高速互聯）理事。通過彼等的前沿見解，王教授及金教授引領我們在技術創新方面位居前列，增強我們提供卓越解決方案的能力，並在這個快速發展的領域保持競爭優勢。

我們的戰略

我們計劃實施以下戰略以進一步發展我們的業務：

不斷提升我們的服務及解決方案組合

AI的快速發展令各行各業出現變革，並重塑雲計算基礎設施的需求，我們持續致力於留在創新的最前沿。我們的戰略立足於對AI技術未來需求的預期，我們致力於提供低時延、開箱即用分佈式雲服務，使智能系統能夠在廣泛應用場景中無縫運行。我們將專注於以下方面，以增強我們的服務及解決方案組合：

- **把握智能體AI激增的機遇，擴大我們的AI雲計算服務規模。**隨著智能體AI的普及，雲計算基礎設施的主要調用者將逐步由人類用戶轉向大規模自主執行複雜、多步驟任務的AI智能體。這一範式轉變將催生全新且持續的雲計算資源需求，其特徵為高頻併發、持久化長上下文記憶及多智能體工作流程編排。智能體AI對算力的此等需求在結構上有別於現有的AI推理工作

業 務

負載，且屬於其之外新增的增量需求。我們已設立智能體基礎設施作為新的業務子分部，並將持續擴大其規模及地區覆蓋、提升多智能體同時運行的速度、穩定性和可靠度。我們亦會推出支持功能，如安全數據存儲、長期記憶體、安全認證使用以及性能追蹤，從而提供一站式智能體基礎設施配置和持續優化。

- *穩步發展我們的邊緣雲計算服務。*我們將繼續完善我們的邊緣傳輸技術，確保提供更高效、更具成本效益的服務。此外，我們將加強我們在算力網絡自主運行方面的專長，使我們能更好地服務於需要高帶寬及低時延交付的應用場景，例如AI視頻生成及低延遲推理服務。
- *升級我們的多模態API平台。*AI技術的進步需要無縫集成文本、音頻、視覺及手勢輸入。因此，我們的目標是建立一個統一的多模態API平台，以連接模型與現實世界的多模態輸入，並支持複雜的智能體工作負載。具體而言，我們計劃開發用於多模態流處理的API支持涵蓋音頻、視覺、手勢及3D空間數據的連續、低時延輸入及輸出流。該等功能對於構建交互式及沉浸式AI系統至關重要。

擴大我們的全球版圖

近年來，全球分佈式雲計算服務市場呈現大幅增長。我們致力把握持續拓展海外市場所帶來的機遇。我們的策略包括進入成熟市場及新興市場以挖掘尚未開發的潛力。具體而言，我們擬開發更多全球計算節點，並與該等地區更多的算力供應商合作，以加強我們基於資源的競爭優勢。

除進行全球拓展外，我們致力大幅提升我們本地服務團隊於主要區域市場的能力。我們將專注於提供定制服務，如銷售、部署、維護及售後支持，以滿足當地客戶獨特且多樣化的需求。這一舉措旨在確保我們的服務種類能夠充分滿足每個市場的獨特需求及期望。

通過利用我們全面的服務組合及憑藉我們強大的研發能力，我們將推出最先進的雲計算解決方案，以緊密貼合當地客戶的特定需求。我們的方法不僅是為了提高用戶滿意度，還致力於建立植根於理解及響應能力的長期關係。通過此承諾，我們將有效地將自己定位為各種國際環境中雲計算服務的首選合作夥伴。

持續提升我們的技術能力

我們將繼續提升我們於雲計算服務行業的技術能力。我們致力於應對日益複雜的AI應用需求，這些應用高度依賴超低延遲、高效且可擴展的解決方案。我們將重點關注以下幾個方面：

- *提高AI推理優化及多模態模型能力。*我們計劃繼續提升大模型推理服務及模型API的性能、成本效益和可擴展性。我們擬以更高算力、更大高速記憶體和更低溝通時延，優化尖端硬件，從而壓降每token推理成本，並提高算力投資總體回報，同時保持服務質量及模型性能。
- *加強綜合智能體基礎設施。*隨著市場更加廣泛採用智能體基礎設施和多智能體協作，我們計劃加強智能體基礎設施，使客戶的智能體工作負載可在我們的平台上以穩定、安全和可擴展的方式運行。我們擬構建不同沙盒實

業 務

例中的一致數據、狀態管理能力，使智能體可在不同場景中維持工作內容。我們亦計劃提供安全認證管理、數據傳輸存儲的端到端加密、自動化備份和數據全生命周期管理，帶來智能體數據的企業級可用性、安全性與合規性。

- **加強AI算力網絡和調度。**我們計劃繼續提高AI算力網絡和調度能力，支持超低時延、大規模、混合工作負載AI應用。我們計劃推行基於服務級別目標層級的差異化調度，此乃按照時延及吞吐量要求將要求分類，並將要求動態發送至AI運算資源的不同層級，從而實現總體利用率最大化，同時保持按需服務質量。與此同時，我們擬繼續開發超低時延算力網絡和雲邊協同，支持邊緣設備處理本地任務，同時將複雜的工作負載轉交雲端處理。

發展並加強我們的社區

我們非常重視我們在全球AI開發者中的聲譽及影響力，因為全球AI開發者是AI社區的重要組成部分。通過新產品、優質的開發者服務及多樣化的增長戰略，我們不斷加強開發者的參與度，提升商業價值並擴大我們的業務範圍。

為加速我們的業務增長，我們成立了專門的客戶獲取及開發者關係團隊，採用新方法來推動業務擴張。我們計劃透過有針對性的數字營銷及社交媒體策略吸引潛在客戶，持續提升產品知名度及品牌影響力。此外，我們將舉辦技術研討會及開發者論壇等線上及現場活動，以維持在社區的強大影響力。我們亦旨在通過積極參與開源社區活動及提供開源集成來吸引開發者使用我們的產品。

同時，我們將與AI社區平台培育新的合作夥伴關係，將我們的產品整合至其生態系統，並支持聯合銷售計劃。我們亦擬與領先的AI應用程序及著名的開源平台合作，以使我們的客戶群多樣化並進一步提高品牌知名度。

在供應方面，我們致力透過拓展多元化供應商渠道及擴大全球合作來加強我們的算力網絡。為滿足多樣化的計算任務需求，我們將與更廣泛的計算資源提供商及其生態系統合作。此外，我們將與全球電信運營商、數據中心、智能計算企業及傳統雲服務提供商建立合作關係。我們將採用按需、現貨定價、收入分成等靈活的定價模式，以適應各種協作需求及業務規模。

我們強大的技術兼容性及靈活的合作模式使我們能夠吸引越來越多的全球算力供應商。這有助我們增強我們的基礎設施，並使我們能夠向客戶提供價格具有競爭力的計算資源。我們相信，該等努力將進一步鞏固我們在分佈式雲計算服務市場的領導地位。

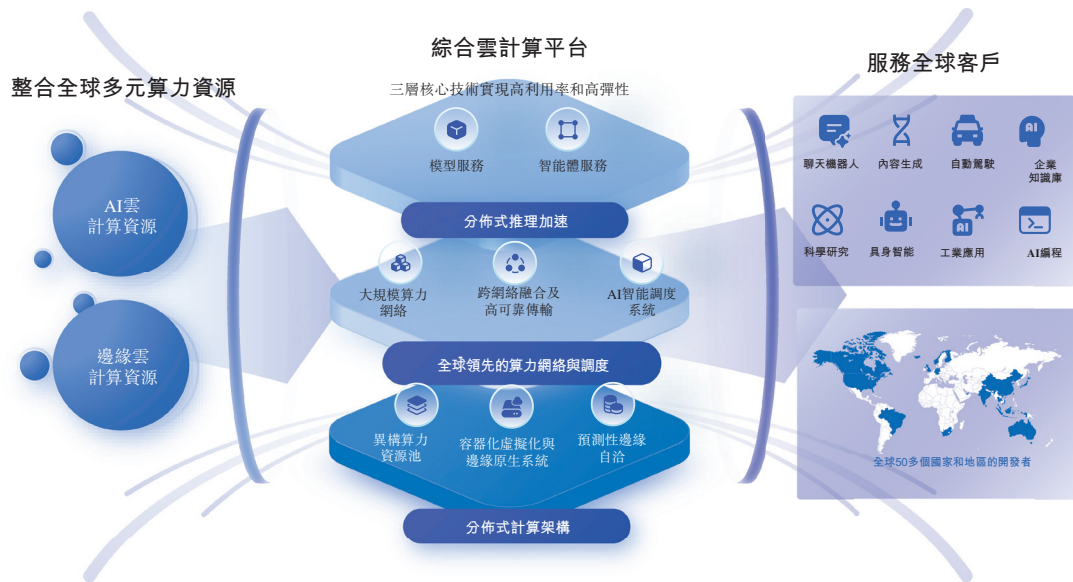
探索戰略聯盟、投資及收購機會

我們計劃有選擇地尋求戰略聯盟、投資及收購，以加強我們的技術能力及鞏固我們的市場地位。著眼於有針對性的擴張及創新，我們的方法包括徹底探索及評估可提供與我們的產品和服務形成互補或產生協同的機會。此舉旨在增強我們的競爭優勢並提升我們在分佈式雲計算服務行業的領導地位。我們相信，該等戰略舉措可提高我們提供前沿解決方案的能力及維持我們的行業領先地位。截至最後實際可行日期，我們尚未確定任何可收購或投資的潛在目標公司。

業務

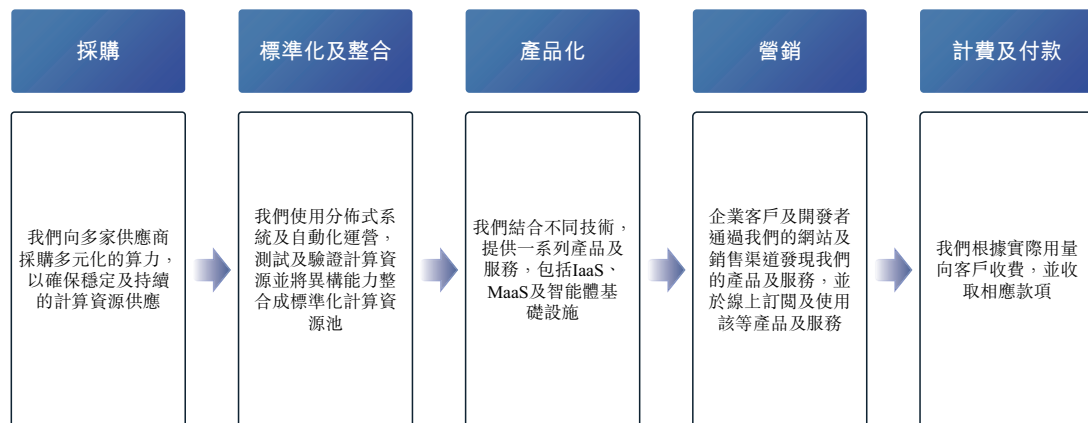
我們的服務及解決方案

我們的業務模式



我們已經建立一個綜合雲計算平台，通過三個集成組件連接全球計算供應與人工智能驅動的需求。在供應方面，我們匯集來自世界各地的兩類計算資源：AI雲端計算資源（包括適用於推理及高效能計算工作負載的分散式GPU容量）；及邊緣雲計算資源（來自地理位置分散的邊緣節點的大規模網絡）。這些異構資源被引導至我們的雲計算平台，該平台跨越三個技術層運行：專有的分佈式計算架構，實現對多樣化硬件的標準化訪問和統一調度；卓越的算力網絡和調度層，通過智能調峰填谷實現資源利用率最大化；及分佈式推理加速層，基於異構計算資源優化模型性能。在需求方面，我們的平台為全球50多個國家和地區的開發者提供服務，支持企業客戶廣泛的應用場景。總而言之，我們通過高效整合全球分散的異構計算資源，並將其轉化為可靠、低成本及高度可擴展的雲計算服務，為不斷增長的全球客戶群創造價值。

下圖載列我們的主要運營工作流程：



業 務

尤其是，我們的服務包括邊緣雲計算服務及AI雲計算服務：

- **邊緣雲計算服務。**我們的邊緣雲計算服務整合第三方提供商的分散異構計算資源，滿足客戶對邊緣計算及傳輸的需求。我們的邊緣雲計算服務並非將數據傳輸至集中式雲數據中心，而是將算力下沉到互聯網「邊緣」（即靠近數據生成及使用的位置）處理。此功能對於需要實時高速處理海量數據的客戶（如音視頻內容提供商）至關重要，乃由於低時延是良好用戶體驗的關鍵。具體而言，我們提供以下服務：
 - **邊緣節點服務。**利用我們廣泛分佈的邊緣計算節點網絡，邊緣節點服務將算力從中心節點遷移至更靠近用戶的邊緣節點，旨在支持邊緣存儲、音頻及視頻處理以及視頻轉碼及分佈等場景。具體而言，我們的產品包括邊緣容器及邊緣裸金屬。邊緣容器是部署在邊緣的虛擬化單元，允許應用程序在數據生成位置或附近高效運行，從而降低時延並提高實時應用的性能。邊緣裸金屬是指部署在邊緣的獨立物理服務器，為計算提供專用硬件資源。
 - **邊緣內容分發網絡（邊緣CDN）。**我們的邊緣CDN依託廣泛的邊緣節點網絡，並利用我們在智能調度、分佈式緩存及網絡傳輸技術方面的專業知識。其主要應用場景包括文件下載、視頻點播及直播。不同於傳統CDN，我們的邊緣CDN與本地網絡深度集成，使服務器更靠近終端用戶。這種近距離部署提升了傳輸質量，為用戶帶來更流暢的視頻播放及更快速的內容分發體驗。
- **AI雲計算服務。**利用集成在我們算力網絡中的分佈式GPU計算資源，我們於2023年推出了AI雲計算服務。隨著AI模型的規模和複雜性不斷提高，我們全面的產品組合滿足了AI推理領域的廣泛需求。這包括提供大型企業所需的彈性GPU計算資源，以及為中小型企業(SME)和開發者量身定制的多樣化模型服務。我們的服務旨在克服全球客戶在採用新技術方面所面臨的成本及技術壁壘。具體而言，我們提供以下涵蓋從基礎到高階功能的產品：
 - **GPU雲服務。**我們的GPU雲服務為客戶提供強大的計算資源，無需購買物理GPU或管理GPU服務器。該架構具有高度彈性，可按需自動擴縮容，採用按需付費模式，客戶只需按實際使用量付費。我們的服務不僅整合了GPU資源，更提供專為生成式AI和機器學習工作負載設計的強大AI運營環境。我們GPU雲服務的主要特性包括優化的推理引擎、預置模板快速部署、需求驅動的自動擴縮容以及微調工具。這些特性使我們的GPU雲服務成為AI推理的全方位平台，為開發和擴展AI應用程序提供高效性、靈活性及可靠性。

業 務

- **模型API**。我們的模型API是客戶訪問預構建、開箱即用模型（主要是LLM及多模態模型）的網關，專為自身缺乏構建、訓練或管理高度複雜AI模型能力的客戶設計。我們提供兩類服務：其一，我們的模型API直接接入領先開源模型，客戶可直接便捷地調用這些開源模型；其二，對於傾向使用自主開發模型的客戶，我們可將有關模型託管在我們的基礎設施上。這使我們的客戶能夠保留對其數據及模型行為的控制權，同時享受我們基礎設施的可擴展性、高效性與可靠性。
- **智能體基礎設施**。AI智能體是透過推理、規劃及採取行動來自主完成任務的AI系統，無需持續的人類監督。智能體基礎設施是一套雲端基礎設施服務，旨在為AI智能體提供可自主運行的雲計算環境，反映出日後AI智能體（而非人類）將成為雲服務主要使用者的趨勢。此項全面技術系統涵蓋AI編程、沙箱執行環境、廣泛的工具與技能，以及短期與長期記憶系統。

下表載列我們於往績記錄期按產品／解決方案類別劃分的收入明細。

	截至12月31日止年度					
	2023年		2024年		2025年	
	金額	%	金額	%	金額	%
(人民幣千元，百分比除外)						
邊緣雲計算服務	358,120	99.9	547,650	98.1	651,128	84.5
－ 邊緣節點服務	313,235	87.4	390,951	70.0	502,645	65.2
－ 邊緣CDN	44,885	12.5	156,699	28.1	148,483	19.3
AI雲計算服務	265	0.1	10,387	1.9	119,217	15.5
總計	<u>358,385</u>	<u>100.0</u>	<u>558,037</u>	<u>100.0</u>	<u>770,345</u>	<u>100.0</u>

邊緣雲計算服務

我們的邊緣雲計算服務並非將數據傳輸至集中式雲數據中心，而是將算力下沉到互聯網「邊緣」（即靠近數據生成及使用的位置）處理。此功能對於需要實時高速處理海量數據的客戶（如音視頻內容提供商）至關重要，因為低時延是良好用戶體驗的關鍵。

邊緣節點服務

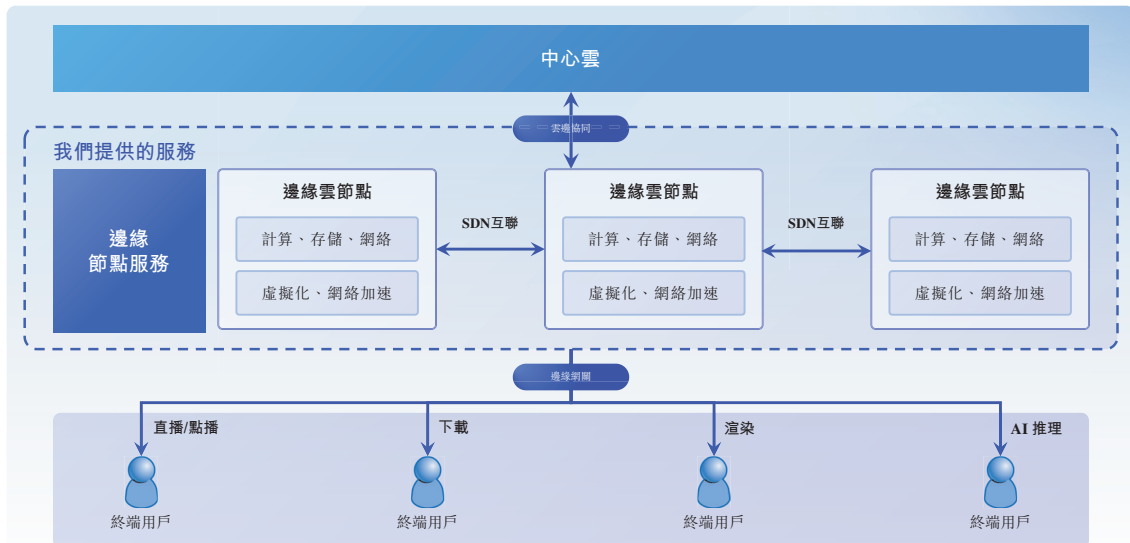
利用我們廣泛分佈的邊緣計算節點網絡，邊緣節點服務將算力從中心節點遷移至更靠近用戶的邊緣節點，旨在支持邊緣存儲、視頻點播、直播以及視頻轉碼及分佈等場景。具體而言，我們的產品包括邊緣容器及邊緣裸金屬：

- **邊緣容器**。「容器」是虛擬化單元，使開發者可將其應用程序打包成單一可移植單元。邊緣容器的功能與傳統容器類似，但部署在邊緣設備或節點上，讓應用程序在數據生成位置或附近高效運行，從而降低時延並提高即時應用的性能。

業 務

- **邊緣裸金屬。**邊緣裸金屬是指部署在邊緣的獨立物理服務器，為計算提供專用硬件資源。與多個用戶或進程共享同一物理機器的虛擬化環境不同，邊緣裸金屬提供對基礎設施的直接、獨佔式訪問，從而實現高性能、低資源開銷及精確控制。

下圖為我們邊緣節點服務的簡要示意圖：



音視頻技術的快速發展及應用場景的不斷拓展，帶動流暢、高清、低時延的媒體體驗需求持續增長。企業如今需要更可靠及更高效的計算資源。憑藉雲邊一體化異構資源管理技術，我們的邊緣節點服務為邊緣容器及邊緣裸金屬提供安全資源隔離、高資源利用率及彈性擴展等核心功能。我們邊緣節點服務的主要特點及優勢包括：

- **異構資源管理。**我們採用先進的異構資源管理技術，可直接訪問物理邊緣裸金屬，解決了虛擬化通常造成的性能損失這一問題，從而實現最佳硬件利用率。我們的服務可滿足計算密集型工作負載需求，確保在保持用戶可訪問性的同時提供最優計算性能。
- **邊緣容器管理。**我們的邊緣容器管理平台支持進程、網絡和文件系統級別的環境隔離、動態資源調度、秒級彈性（即系統可自動秒速擴縮容，以應付對計算資源的需求波動）及快速業務部署。
- **全面智能監控系統。**強大的全天候監控框架支撐我們的邊緣節點服務，通過應用機器學習提供故障預測及異常信息預警，確保主動檢測及解決問題，保障運營可靠性。

案例：對一家領先個人雲存儲平台的邊緣節點服務

我們的客戶是中國領先的個人雲存儲服務商，擁有海量活躍用戶。隨著用戶生成內容的快速增長，平台每天需處理PB級的非結構化數據。

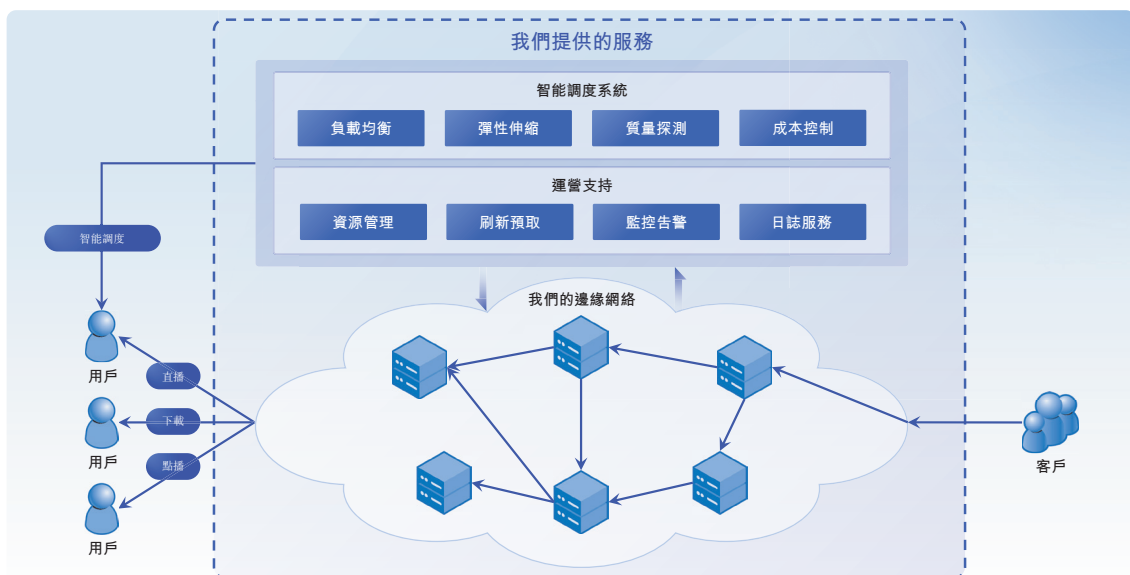
業 務

- **客戶的挑戰。**我們的客戶面臨巨大的性能及成本壓力。在早晚等高峰時段，集中訪問熱門文件常導致下載速度驟降，尤其是大文件傳輸時，導致用戶體驗較差。集中式存儲架構亦推高帶寬成本（佔運營開支的35%以上）。
- **我們的解決方案。**針對這些問題，我們提供邊緣節點服務賦能的智能存儲解決方案：(a)構建覆蓋全國邊緣節點的分佈式緩存網絡。緩存是一種將經常讀取的數據臨時儲存在比原始數據源更快可檢索位置的技術。我們開發智能算法，將熱門內容預緩存在更靠近用戶的邊緣節點上，並提升下載速度；(b)分佈式緩存網絡提供更大存儲容量及提升緩存命中率。緩存命中率是衡量緩存工作效率的指標。其代表在緩存中找到所要求數據的次數與發出數據要求總數的比例（「緩存命中率」）。透過提升緩存命中率，我們可提高下載速度和降低算力資源成本；及(c)應用加速技術，優化在弱網環境下的下載速度，進一步提升用戶體驗。

我們的解決方案為客戶帶來顯著效益。即使在弱網環境下，也能實現更快、更流暢的下載，從而提升數億月活躍用戶的體驗。優化後的帶寬使用方案使運營成本降低近30%。此外，增強的數據安全性及高服務可用性助力客戶提升用戶信任度。

邊緣CDN

我們的邊緣CDN由廣泛的邊緣節點網絡組成，並利用我們在智能調度、分佈式緩存及網絡傳輸技術方面的專業知識。其主要應用場景包括文件下載、視頻點播及直播。不同於傳統CDN，我們的邊緣CDN與本地網絡深度集成，使服務器更靠近終端用戶。這種近距離部署提升了傳輸質量，為用戶帶來更流暢的視頻播放及更快速的內容分發體驗。下圖為我們邊緣CDN的簡要示意圖：



業 務

我們的邊緣CDN具有以下主要特點及優勢：

- **智能調度。**我們的智能調度機制以毫秒級速度分配及調度計算節點。其管理數萬台設備的運行狀態及文件緩存分佈，可實時確定距離用戶最近的節點。如某一節點出現故障，我們的系統可快速將工作負載轉移至附近的另一節點，確保用戶請求不受影響。
- **彈性計算。**利用「容器化」技術，我們實現快速且高度靈活的節點部署，可實時調整分佈。這種設置確保了服務的高可擴展性，以應付如電商平台在全國促銷活動期間遇到高併發請求的情況。
- **成本控制算法。**我們的成本控制算法基於靈活的節點管理及調度系統運行。該算法通過分析歷史趨勢預測帶寬冗餘，力求最大限度利用網絡容量，盡量減少閒置算力。

案例：對中國一家知名短視頻平台的邊緣CDN服務

我們的客戶是中國一家知名短視頻平台，目前正處於快速發展階段。隨著觀看流量的攀升，平台面臨使用量突增及內容分發效率不足的雙重挑戰。

- **客戶的挑戰。**我們客戶的平台在明星直播期間會面臨區域性流量激增，此類大型直播活動因流量集中在特定區域而造成網絡擁堵。此外，偏遠低線城市及縣域用戶飽受播放緩沖時間過長的困擾，嚴重影響使用體驗。傳統的CDN架構限制了平台有效平衡用戶體驗和成本效益的能力，導致平台難以管理流量高峰，同時面臨高昂的運營開支。
- **我們的解決方案。**我們的邊緣CDN服務採用智能算力調度，建立動態負載平衡機制。該方案自動檢測不同計算節點之間的工作負載不均衡，實時重新路由流量，從而優化網絡性能。

此外，我們的邊緣節點亦部署在低線城市及縣域，有助優化「最後一英里」的內容傳輸，提升整體內容分發效率。我們的平台亦支持帶寬資源複用，利用閒時網絡資源預加載高訪問內容，從而創建靈活的資源池，提高內容可及性。

AI雲計算服務

我們的AI雲計算服務為AI計算提供高性價比算力。通過我們的分佈式算力網絡，我們整合全球的GPU資源。這使我們能夠使用靈活的按使用量收費定價模式為AI微調和大規模AI推理任務提供計算資源。此外，通過整合一系列廣受認可的開源LLM和多模態模型，並進一步優化推理過程，我們為全球的企業和AI開發者提供基於API的服務。這種服務幫助客戶減少採用AI技術的相關壁壘和成本。

業 務

GPU雲服務

圖形處理單元(GPU)擅長處理需要大量算力的並行處理任務。通過使用我們的GPU雲服務，客戶可訪問強大的計算資源，無需自行購買物理GPU或管理GPU服務器。我們的架構具備高度彈性的可擴展性，可按需自動擴縮容，且客戶只需按實際算力用量付費。此外，我們的服務並非GPU資源的簡單整合。我們優化AI模型推理的運行環境，使推理任務的處理速度及效率更高。此外，客戶亦可在我們的平台輕鬆部署模型。

我們的GPU雲服務主要應用於兩類應用場景：AI微調和大規模AI推理。在AI微調方面，我們通過按使用量收費定價模式，向客戶提供豐富的GPU資源。在大規模AI推理方面，我們提供高度可擴展的分佈式GPU資源，以及對多樣化AI模型的訪問權限。

我們的GPU雲服務具有以下主要特點及優勢：

- **彈性可擴展性。**我們的架構可根據需求自動擴縮容，我們的客戶只需按算力的實際使用量付費。此功能對需在高峰需求期間快速擴展的客戶（如電子商務企業）而言至關重要。
- **優化運行環境。**我們的GPU雲服務並非GPU資源的簡單整合。我們優化AI模型推理的運行環境，使推理任務的處理速度及效率更高。
- **按使用量收費(*pay-as-you-go*)。**我們的GPU雲服務採用按使用量收費定價機制。該機制通過提供與實際使用量一致的彈性定價策略，巧妙地滿足客戶在關鍵需求期間靈活分配資源的需求。其為客戶提供具有成本效益的解決方案，其中支出與實際使用的資源相對應。
- **大型模型高速下載。**通過為海量交付提供先進的網絡能力，我們的GPU雲服務大幅減少了等待時間，使用戶能夠快速訪問大型模型及數據集。這提高了用戶的生產力及效率，使其更容易在有限的時間內處理複雜的任務。
- **容器化解決方案。**此方案可確保計算環境的有效隔離，從而防止對資源的競爭或依賴項的衝突。我們的GPU雲服務亦提供用戶友好的操作界面，從而通過簡化的工作流程提高生產力。
- **共享存儲。**該解決方案提供高性能的共享雲存儲設施，有利於跨多個GPU雲服務無縫共享數據，確保全部相關用戶訪問一致的數據。該功能對於加強協作以及促進有效的協調及數據管理實踐至關重要。

案例：為一家領先的汽車軟件公司提供GPU雲計算服務

客戶是著名的智能座艙軟件解決方案提供商，為全球中高端汽車製造商提供全面的智能汽車軟件及數字化服務解決方案。

業 務

- *客戶的挑戰*。客戶的業務需要對大模型進行微調，而其最初依賴的是閉源模型。然而，由於早期的閉源模型的版本迭代有限，故這方法的成本高昂。儘管其後客戶轉用開源模型，但是微調期間的GPU租賃成本仍然很高，計算開支佔預算的重大部分。GPU資源能否獲得長期穩定的供應仍是一個令人關切的問題。

部署後，微調過的模型在處理推理服務方面也有挑戰。流量在高峰時段與非高峰時段之間劇烈波動，流量激增會給資源分配帶來重大挑戰，使提供無縫的用戶體驗變得更加困難。

- *我們的解決方案*。(i)在對開源模型及閉源模型的成本效益及性能進行仔細比較後，我們協助客戶過渡到開源模型，從而降低其成本並穩定訓練數據輸出。我們為LLM實施統一的微調及推理工具包，使客戶能夠使用同時支持基礎模型及微調模型的單個實例有效地驗證微調的有效性；
 - (ii) 我們建議適合客戶需求的高性能GPU，並提供可靠的GPU運算資源，以確保模型的穩定性及持續的優化；及
 - (iii) 憑藉我們在分佈式計算及算力調度方面的專業知識，我們向客戶提供按使用量收費(pay-as-you-go)的定價，從而能夠根據GPU使用量進行精確計費。在流量高峰期，資源自動增加，而在非高峰期利用具有成本效益的計算資源減少開支。

我們的解決方案為客戶帶來實質收益。過渡到開源模型降低了培訓成本並提高了數據一致性。優化的GPU供應確保了穩定的長期供應，支持穩定的模型迭代及微調。彈性計算基礎設施可在需求激增期間提供可擴展的資源，同時在非高峰期有效降低運營成本。這種綜合方法不僅提高了服務績效及增強了用戶滿意度，還提升了客戶在汽車軟件行業的競爭地位。

案例：為建築軟件領域的全球領導者提供GPU渲染服務

客戶是一家業界領先的建築軟件公司，擁有先進的實時渲染算法。為了提高設計效率及增強創意體驗，彼等需要高度優化的實時渲染能力，這對算力調度及推理性能提出了很高的要求。

- *客戶的挑戰*。該客戶遇到調度服務效率低及推理服務時延高等問題，從而影響了實時渲染的效率及流暢性。此外，依賴公有雲網絡進行數據傳輸也會存在安全隱患，亦導致成本高昂。
- *我們的解決方案*。我們提供全面的GPU/CPU容器化服務，實現渲染工作負載的無縫部署及管理。通過創建安全的虛擬私有雲(VPC)，我們確保CPU及GPU實例之間快速可靠的內部數據傳輸，消除直接使用公共網絡的風險，降低時延並增強安全性。高性能共享雲存儲支持高效的多實例數據訪

業 務

問，簡化工作流程。為確保可靠性，我們一週七天一天二十四小時的運維支持及時處理問題及保持平台穩定，以一致的快速回應性能幫助客戶滿足其實時渲染要求。

我們的解決方案顯著降低了網絡時延並優化了運營成本。憑藉我們在AI基礎設施方面的支持，該客戶得以專注於提升其核心渲染能力，從而鞏固彼等在建築軟件創新方面的領導地位。

模型API

我們的模型API為預構建模型，可迎合需要AI技術但缺乏自行構建、訓練及管理高度複雜AI模型能力的客戶。我們的模型API包括兩大類：大語言模型及多模態模型。大語言模型處理各種自然語言處理任務，而多模態模型處理並生成圖像及視頻。我們提供兩種類型的服務：首先，我們的模型API提供對領先開源模型的直接訪問，使客戶能夠直接便捷地使用該等開源模型；其次，對於更偏向使用其自行開發模型的客戶，我們在自有基礎設施上託管此類模型。這使我們的客戶能夠保留對其數據及模型行為的控制權，同時受益於我們基礎設施的可擴展性、效率及可靠性。下圖是我們模型API的簡化圖：



我們的模型API具有以下特點及優勢：

- **快速集成**。我們的模型API提供跨多種模態的最新開源模型的廣泛訪問權限型。這種簡化的方法允許企業直接部署先進AI工具，節省傳統模型搭建通常所需的時間。
- **維護成本低**。我們的模型API使客戶擺脫複雜的模型部署及運維工作。我們的開箱即用型服務旨在處理模型維護過程中的複雜問題，以便我們的客戶能夠專注於實現業務目標而不分心。
- **成本效益**。我們的模型API採用按使用量收費(pay-as-you-go)的定價模式，與客戶的實際使用情況保持一致。該定價模式由穩健的分佈式彈性調度框架支持，可進一步優化性能及成本效益。
- **模型種類多樣**。我們的平台擁有動態的模型組合，能夠在廣泛的生態系統中實現高效加載與靈活切換。這種靈活性消除了單個客戶管理模型的需要，從而提供了一個隨客戶需求而發展的適應性平台。

業 務

案例：與全球領先的AI角色扮演遊戲公司集成Model API

客戶為全球領先的AI動漫角色扮演遊戲開發商。客戶擁有一支中型研發團隊，具有紮實的模型優化能力。

- **客戶的挑戰。**客戶需要維護自己的GPU服務器，同時確保其網絡基礎設施的穩定性。在流量增加期間，用戶響應時間往往會延遲。
- **我們的解決方案。**憑藉我們開箱即用的模型API，客戶可便捷地將模型與其業務需求保持一致，並可不時進一步優化模型，以應對其不斷變化的需求及偏好。通過使用我們的模型API服務，我們大幅降低了客戶的基礎設施維護成本。按使用量收費(pay-as-you-go)的定價模式亦有助客戶降低整體開支。我們的模型API亦包含自動擴展機制，旨在解決高峰流量期間響應時間增加的問題。該等未使用的GPU資源將重新用於模型改進及訓練，從而進一步提高客戶的成本效益。

案例：為AI助手創新開發者提供全球一站式LLM API服務

客戶是一家領先的AI創新公司，提供一站式AI助手，提供聊天、搜索、協作及編程支持等功能。隨著AI技術的迅速發展，客戶需要通過將前沿的AI工具及功能集成到其產品線中來保持領先地位。

- **客戶的挑戰。**客戶在快速驗證AI新功能的可行性及有效性以調整其發展戰略及避免資源浪費方面面臨重大挑戰。這需要訪問穩定可靠的LLM API，但許多可用的API缺乏必要的一致性、可靠性及成本效益。
- **我們的解決方案。**利用我們專為大規模AI應用而設計的優化分佈式推理和推理加速功能。我們的高可靠性、可擴展且具有成本效益的LLM API服務使客戶能夠迅速構建及驗證生成式AI應用程序。我們支持的模型在角色扮演、故事擴展、代碼生成及同理心互動等方面表現出色，將高智商與情感共鳴相結合。對於編程支持，我們的DeepSeek R1 API發揮了特別重要的作用。其支持函數調用能力，賦能客戶高效地測試AI特性並加速開發週期。

智能體基礎設施

AI智能體是透過推理、規劃及採取行動來自主完成任務的AI系統，無需持續的人類監督。智能體基礎設施是一套雲端基礎設施服務，旨在為AI智能體提供可自主運行的雲計算環境，反映出日後AI智能體（而非人類）將成為雲服務主要使用者的趨勢。此項全面技術系統涵蓋AI編程、沙箱執行環境、廣泛的工具與技能，以及短期與長期記憶系統。

我們提供兩大類能力。首先，我們的智能體基礎設施為AI生成編碼、工具調用及瀏覽器操作提供安全、隔離的執行環境。這些沙箱環境防止AI智能存取或篡改沙箱外的資源，確保智能體的行為不會對客戶的生產系統造成損害。這些沙箱環境支持運行環境快速啟動、暫停及恢復以及複製，有助客戶靈活管理長期運行、間歇或高度並行的智能體工作負載，並壓降閒置期間的資源成本。第二，我們的智能體基礎設施支持智能體之間的託管和編排。通過標準化界面和開發者工具，客戶可配置本身的智能體裝置作為服務、按需管理運營並擴大其能力，同時免受相關基礎設施複雜性的影響。

業 務

我們的智能體基礎設施具有以下特色和優勢：

- **安全的操作環境。**硬體虛擬化及沙箱技術，對智能體工作負載構建強大的隔離邊界，確保AI生成編碼和工具調用在受控環境中運行，而不影響其他裝備或系統。
- **全面操作狀態管理。**完美支持暫停、恢復、快拍及複製，支持智能體長時間保存內容，閒置時處於休眠狀態，但可按需恢復，提高長時間運行或間歇性任務的效能。
- **彈性的可擴展性。**我們的基礎設施可按任務容量及複雜性，自動增減智能體執行環境規模，此對流量波動或大規模智能體同步交互的場景尤其重要。
- **便利開發者的體驗。**標準化軟件開發包及配置工作，簡化智能體應用的開發和操作，支持客戶毋須重建環境，即可快速從算法實驗進入生產配置。

用例：AI生成編碼智能體基礎設施

我們的客戶為AI智能體開發者，其AI生成編碼工作負載在我們的基礎設施上運行。

- **客戶的難題。**在AI生成編碼場景中，沙箱環境通常在等待模型輸出或用戶輸入期間保持活躍（即使進行小量實際運算亦然）。客戶在配置其智能體時，亦須解決並行程度較高之時，管理容量和基礎設施開支的操作複雜性。客戶努力在較輕量的操作環境達致彈性擴展及推出智能體服務，同時利用我們的智能體基礎設施提效減費。
- **我們的解決方案。**我們以智能體沙箱，為客戶提供虛擬化的隔離運行環境，解決以上難題。運行環境可在等待模型輸出或用戶確認期間，在數秒之內暫停及恢復，如此大幅壓降長期間置期間的相關資源成本。沙箱亦可按照客戶的最新編碼，快速創建全新環境，確保用戶均可在最新的隔離工作環境工作，同時緩減共享環境的相關操作風險。以上多種能力，使我們的智能體基礎設施能夠支持客戶以節省成本、穩健操作的方式，運行大量AI生成編碼工作負載。

業務可持續性及盈利路徑

我們的歷史業績

於往績記錄期，我們的收入由2023年的人民幣358.4百萬元迅速增長至2025年的人民幣770.3百萬元，2023年至2025年的複合年增長率為46.6%。特別是，自2025年起，我們的AI雲計算服務出現爆炸式增長，收入由2024年的人民幣10.4百萬元增加逾十倍至2025年的人民幣119.2百萬元。我們的毛利由2023年的人民幣63.5百萬元迅速增加至2024年的人民幣68.8百萬元，並進一步增加至2025年的人民幣72.0百萬元，2023年至2025年的複合年增長率為6.5%。我們的毛利率由2023年的17.7%下降至2024年的12.3%，並進一步下跌至2025年的9.4%。

業 務

儘管我們實現了持續業務增長，但我們於2023年、2024年及2025年產生的年內虧損分別為人民幣189.4百萬元、人民幣293.5百萬元及人民幣222.9百萬元，同年經調整淨虧損（非香港財務報告準則計量，不包括可轉換可贖回優先股及股份支付費用的公允價值虧損）分別為人民幣37.1百萬元、人民幣61.6百萬元及人民幣104.6百萬元。由於我們於2023年推出AI雲計算服務，該業務分部於往績記錄期正處於早期發展階段，尚未實現盈利。就我們的邊緣雲計算服務業務而言，我們於整個往績記錄期維持毛利，且我們預期於往績記錄期對研發及相關基礎設施的投資將推動該業務分部的可持續增長。

於往績記錄期，我們的淨虧損主要歸因於以下因素：

主要與AI雲計算服務相關的因素：

- *AI雲計算服務處於增長階段，尚未達到規模經濟。*我們於2023年推出AI雲服務。該業務分部於往績記錄期處於早期發展階段，其收入尚未達到可讓我們實現規模經濟的水平。於2023年、2024年及2025年，我們來自AI雲計算服務的收入分別為人民幣0.3百萬元、人民幣10.4百萬元及人民幣119.2百萬元，分別佔我們同期總收入的0.1%、1.9%及15.5%。與我們較為成熟的邊緣雲計算服務相比，AI雲計算服務的計算節點數量及付費客戶數目仍在不斷增加。然而，為確保向客戶提供低延遲、高可用性及具彈性的容量，我們已需要提前預留具一定水平的計算資源，並維持可支援大規模推理工作負載的分佈式基礎設施。無論實際使用量的短期波動如何，均會產生該等預留資源及基礎設施成本。因此，我們AI雲計算服務的單位收入成本仍然相對較高，而我們的收入尚不足以充分吸收該等預留容量成本，導致該業務分部於往績記錄期出現淨虧損。
- *用於建設AI雲基礎設施的重大投資。*我們為建立AI雲計算服務而產生巨額前期投資。截至2024年及2025年12月31日，我們的AI雲計算服務下的計算節點數量分別為19個及64個，而截至2026年4月30日則增加至77個。我們的前期投資包括採購及集成分佈式計算資源、建設及配置服務器及相關硬件，以及搭建支援大規模AI推理工作負載所需的基礎架構。隨著我們計算節點的增加，AI雲計算服務的計算資源成本快速增長。於2023年、2024年及2025年，我們產生的計算資源成本分別為人民幣0.2百萬元、人民幣9.4百萬元及人民幣117.4百萬元，分別佔我們同年AI雲計算服務銷售成本的29.8%、46.5%及89.0%。

此外，建立我們的AI雲計算服務需要我們投資於專用硬件。該等資產一經投入使用，我們即開始按其估計可使用年期確認折舊，導致初始期間的折舊相對較高，即使來自AI雲計算服務的相關收入仍處於早期增長階段。於2023年、2024年及2025年，我們產生的折舊成本分別為零、人民幣7.0百萬元及人民幣5.3百萬元，分別佔我們同年AI雲計算服務銷售成本的零、34.4%及4.0%。

業 務

- **龐大研發開支。**我們錄得龐大研發開支，以迭代及提升我們的AI雲計算服務，當中明確重點投放於我們預期長遠將改善盈利能力的技術。例如，我們推進有關分佈式推理加速技術的研究，該技術旨在為客戶實現更快的AI推理及優化我們內部計算資源的利用率。該研究涉及在實際場景中進行廣泛的測試，並輸入大量數據，以驗證技術的有效性、穩定性及可靠性。該測試屬計算高度密集型，需要大量的GPU計算資源及雲端服務，進而產生了研發開支。於2024年至2025年，我們的測試開支由人民幣2.4百萬元大幅增加至人民幣24.3百萬元，分別佔研發開支的2.8%及20.3%。測試開支的快速增長反映我們持續投資於突破AI推理加速技術可實現的上限，以及嚴格驗證雲計算服務與新推出的AI模型的兼容性。

此外，開發AI雲計算服務需要高技能的研發人員，彼等一般享有具競爭力的薪酬水平。於往績記錄期，我們持續擴大研發團隊，尤其是專注於AI相關技術（如大型模型優化及推理引擎開發）的人員，以提升我們的研發能力並為客戶提供更新的服務及功能。於2023年、2024年及2025年，我們產生的僱員薪金及其他福利分別為人民幣54.9百萬元、人民幣71.6百萬元及人民幣80.8百萬元，分別佔我們同期研發開支的80.9%、82.9%及67.3%。該等開支反映了我們研發團隊規模的增長以及經驗豐富的AI工程師及研究人員的平均薪酬水平不斷上升。隨著我們進一步開發AI雲計算服務並致力維持技術領先地位，我們預期將繼續產生巨額研發人員成本。

與邊緣雲計算服務及AI雲計算服務相關的因素：

- **運行及維護邊緣雲及AI雲計算基礎設施的持續開支。**除上文所述建設雲計算基礎設施的前期投資外，我們亦產生運行及維護該計算基礎設施的經常性成本，包括計算資源成本（如服務器成本、電力及維護成本）及雲服務開支。運行AI工作負載（尤其是大規模AI推理）屬耗電及計算密集型工作，需要我們維持充足的容量以滿足不同地區客戶對延遲及可用性的要求。此外，我們持續產生有關數據儲存、安全及系統監控的開支，以及為緊貼AI技術發展而定期升級或更換硬件的相關開支。

我們的盈利路徑

利用AI的激增擴展規模經濟

目前，AI推理已取代AI訓練成為算力的主要需求來源。AI推理應用大規模滲透至消費者及企業的工作流程中，對高效、彈性且具成本效益的推理服務產生了持續且高頻的需求。展望未來，隨著智能體AI的普及，雲計算基礎設施的主要調用者將逐步由人類用戶轉向大規模自主執行複雜、多步驟任務的AI智能體。這一範式轉變將催生全新且持續的雲計算資源需求，其特徵為高頻併發、持久化長上下文記憶及多智能體工作流程編排。智能體AI對算力的此等需求在結構上有別於現有的AI推理工作負載，且屬於其之外新增的增量需求。

業 務

我們已設立智能體基礎設施作為新業務子分部，並將持續升級及擴充該子分部的產品形態。例如，我們將進一步升級我們的智能體沙箱，該產品為執行AI生成的編程碼提供一個安全、隔離的環境。該沙箱環境可防止AI智能體存取或篡改沙箱外的資源，確保AI智能體的行為不會對更廣泛的系統造成危害。

隨著智能體AI時代的展開，智能體基礎設施的使用將成規模化，而我們預期規模經濟將進一步擴大：更大及更多元化的智能體工作負載基礎將產生流經我們全球計算調度網絡的額外token消耗量，從而提高整體資源利用率，降低單位計算成本並進一步加強我們AI雲計算服務的盈利能力。

與開源社區合作，推行開發者友好的增長戰略

我們與擁有廣泛開發者網絡的開源社區（如GitHub及OpenRouter）合作。透過將我們的產品整合至合作夥伴生態系統並推行聯合市場推廣計劃，我們預期可加速獲取新客戶、擴大行業覆蓋範圍並提高每名客戶收入。於2026年1月，我們在OpenRouter的第三方開源大模型提供商中排名第一，並於2025年12月在Hugging Face的推理服務月請求量排名第一。

我們亦實施開發者友好戰略，以相對較低的獲客成本吸引新客戶。我們積極參與開源社區活動，提供開源整合，並舉辦線上及線下技術研討會及開發者論壇。於2025年，我們的AI雲計算服務擁有超過180,000名活躍用戶。我們計劃透過將更多開發者轉化為付費客戶、推廣更高價值服務以及鼓勵其所在機構擴大使用範圍，進一步發掘該用戶群體。

拓寬我們的客戶觸達及擴展我們的服務網絡

隨著我們擴展服務網絡及擴大客戶基礎，我們致力於利用我們的領導地位，把握分佈式雲服務市場的巨大增長機會。我們擬通過擴大我們的全球影響力、培養與不同行業主要客戶的關係及建立一個繁榮的開發者生態系統來推動持續增長。這些舉措加上高效的營運及規模經濟效應，使我們能夠實現長期盈利，同時為客戶創造卓越價值：

- **加強與邊緣雲服務主要客戶的關係。**我們的邊緣雲計算服務主要服務於中國的大型互聯網公司。根據灼識諮詢的資料，我們的客戶包括中國互聯網十強企業中的大部分。我們與該等客戶的長期合作令我們在技術、服務及產品方面實現了深度整合。展望未來，我們將繼續提供優質的客戶支持及全生命週期服務。我們擁有專門的售後及項目管理團隊，可提供快速響應、持續進度更新、技術諮詢及實際實施支持，尤其是對於具有複雜集成需求的客戶。這可提高客戶滿意度和留存率，並為交叉銷售及追加銷售其他服務創造機會。
- **在全球擴展我們的計算資源網絡。**我們在全球範圍內持續擴展我們的計算資源網絡。特別是，就我們的AI雲計算服務而言，我們的計算節點數量由截至2024年12月31日的19個增加至截至2025年12月31日的64個，並進一步增加至截至2026年4月30日的77個。我們重點關注以下主要地區，商業原因載列如下：(i)北美，該地區的GPU資源供應相對充足，客戶對該等計算資源的需求較高；因此，我們正積極擴展我們在北美的計算資源，以配

業 務

合我們於該地區不斷擴大的客戶群；(ii)東南亞。繼在新加坡建立了我們於東南亞的第一個計算節點後，我們進一步在印度尼西亞、泰國及馬來西亞等國家增加計算節點，以利用該等國家電力成本較低的優勢；及(iii)日本（尤其是日本北部地區），氣候條件優越（氣溫較低，散熱能耗較低），且當地政府提供的能源補貼有助計算服務增長。

- 通過客製化解決方案拓展至新的垂直行業。我們將繼續在不同行業擴大我們的客戶基礎。目前，我們主要服務於泛娛樂、電信及信息科技領域的客戶。我們擬深化在現有垂直領域的滲透，並拓展至教育、醫療保健及金融服務等其他行業，這些行業對低時延、高可靠性及具成本效益的算力的需求預計將持續增長。例如，我們正在探索如何利用我們的服務支持電商公司的內容生成工具及營銷活動，使其能夠更高效、更方便地創造內容。同樣，我們正在為教育公司開發AI雲計算服務與AI輔助教學工具之間的整合方案，幫助其以更具成本效益的方式利用AI功能。通過提供該等行業特定解決方案，我們旨在獲取新客戶並增加現有客戶的使用廣度和深度。

升級我們的技術及服務

憑藉我們現有產品及服務的成功，我們計劃進一步優化服務，提升其適應性、準確性及效率，從而增強我們的競爭優勢。

- 首先，我們將繼續升級智能體沙箱，即智能體基礎設施業務分部下的新產品。智能體基礎設施是專門為支持AI智能體自主、持久地執行任務而構建的基礎設施。我們的智能體沙箱是一個安全、隔離的環境，用於執行AI生成的代碼。這個沙箱環境可防止AI智能體訪問或篡改沙箱之外的資源，從而確保AI智能體的行為不會對更廣泛的系統造成損害。我們正在將我們的智能體沙箱升級到一個更強大的環境，該環境不僅可以隔離AI生成的代碼，還提供更強的安全控制並支持更複雜、更長時間運行的智能體工作流程。
- 其次，我們計劃發展超低時延算力網絡。超低時延算力網絡是支撐先進AI系統在自動駕駛、遠程手術、智能製造等場景下實現毫秒級響應速度及實時決策的關鍵基礎設施。
- 第三，我們將通過確保我們的平台支持越來越多樣化的模型來增強我們的模型託管能力。這種適應性對於滿足客戶的不同業務需求及偏好至關重要。
- 第四，我們旨在持續加強雲與邊緣系統的整合。邊緣設備將獨立管理本地任務，並將更複雜的工作負載發送到雲端進行組合處理。這種改進的協調機制可以提高網絡吞吐量並減少端到端延遲。
- 最後，我們擬促進下一代智能硬件（如NPU）與我們的算力網絡集成。此舉將提高我們支持長期運營及應對高併發計算資源請求場景的管理能力。該等舉措不僅豐富我們的服務產品，亦可鞏固我們在快速發展的AI領域的地位。

業 務

優化運營效率

我們管理及優化開支的能力對我們業務的成功及我們實現可持續盈利的能力至關重要。

於往績記錄期，我們產生了大額經營開支，包括研發開支、銷售及營銷開支以及行政開支。下表載列所示年度我們的研發開支、銷售及營銷開支以及行政開支佔收入的百分比：

	截至12月31日止年度		
	2023年	2024年 (%)	2025年
佔收入的百分比：			
研發開支.....	18.9	15.5	15.6
行政開支.....	7.1	5.6	7.4
銷售及營銷開支.....	5.6	4.6	5.5
總計.....	31.6	25.7	28.5

- **研發開支。**於2023年、2024年及2025年，我們的研發開支分別為人民幣67.8百萬元、人民幣86.3百萬元及人民幣119.9百萬元，分別佔我們總收入的18.9%、15.5%及15.6%。我們的研發開支佔收入的百分比總體下降，主要是由於收入快速增長及隨業務增長實現的規模經濟效應。隨著我們持續推動服務迭代及技術升級，我們期望憑藉我們積累的技術能力進一步提升研發效率。
- **行政開支。**於2023年、2024年及2025年，我們的行政開支分別為人民幣25.3百萬元、人民幣31.3百萬元及人民幣57.3百萬元，分別佔我們總收入的7.1%、5.6%及7.4%。撇除[編纂]的影響，我們於2025年的行政開支為人民幣40.4百萬元，佔我們總收入的5.2%。於往績記錄期，我們的行政開支(不包括[編纂])佔收入的百分比持續下降，主要歸因於我們採取有效的成本控制措施及努力提升運營效率。
- **銷售及營銷開支。**於2023年、2024年及2025年，我們的銷售及營銷開支分別為人民幣20.2百萬元、人民幣25.8百萬元及人民幣42.4百萬元，分別佔我們總收入的5.6%、4.6%及5.5%。我們的銷售及營銷開支佔收入的百分比有所減少，主要是由於規模經濟效應導致收入快速增長及我們有能力提升銷售及營銷活動效率。

我們深明，對經營開支進行規範的管控對實現及維持盈利能力至關重要。通過將技術驅動的效率、有效的成本管理及可擴展的運營模式相結合，我們處於有利位置，可提升盈利能力，同時繼續投資於創新、增長及長期成功。

受益於上述收入增長驅動因素及成本控制措施，我們預期我們的收入將持續增長。我們亦致力於擴大業務以實現規模經濟，從而更有效地管理經營開支及支持我們實現盈利。因此，董事認為，且聯席保薦人同意，我們的業務具有可持續性。

業 務

營運資金充足性

經計及我們的可用資源，包括現金及現金等價物、定期存款、已抵押存款、經營現金流量、銀行借款、可用銀行融資及我們可用的[編纂]估計[編纂]，董事認為且聯席保薦人認同，我們擁有足夠的營運資金滿足我們目前以及自本文件日期起至少未來12個月的需求。

我們的核心技術

專有分佈式計算架構

異構算力資源池。我們的系統將多樣化的算力資源集成到一個統一的資源池中。通過智能工作負載編排，我們的架構減少了底層硬件的異構性，實現了多種計算單元的無縫集成，同時為上層應用提供了標準化的訪問方式。

容器化虛擬化及邊緣原生系統。我們的系統基於輕量級的雲原生架構，沿兩個互補層面高效組織碎片化的邊緣資源：一個AI計算層，處理AI推理及高性能計算工作負載；及一個邊緣計算層，通過分佈式邊緣節點網絡推動對延遲敏感的服務。該設計將工作負載根據性能需求、延遲限制及成本目標智能地導向最佳資源層級，同時較傳統解決方案大幅減少系統開支。這種超緊湊的設計支持在單服務器邊緣節點上進行部署，即使環境受限的情況下也能最大限度地利用資源。該系統包括Serverless GPU技術，該技術實現自動擴展、故障恢復和硬件無關的部署。通過使用容器和覆蓋網絡實現統一編排，確保無縫任務遷移和跨地理分散的節點的多租戶隔離。

算力網絡及調度能力

最大的算力網絡。截至2025年12月31日，我們已建立了覆蓋逾1,340個縣市的算力網絡，整合全球逾4,600個計算節點。這種高密度計算節點部署可實現毫秒級時延（短至10毫秒）本地分佈，為實時視聽內容分發及AI推理等場景提供基礎設施支持。

全球調度能力。我們的全球調度網絡使我們能夠持續、全天候地實現計算資源的高利用率。通過利用跨地區時區差異和我們多樣化客戶群的交錯需求概況，我們將全球閒置的計算能力動態分配至任何特定時刻的需求高點，確保我們的GPU資源得到有效部署而不是在非高峰時段閒置。因此，我們的平均GPU利用率於2025年一直高於75%，根據灼識諮詢，大幅超越行業平均水平40%至50%。這種智能的「削峰填谷」能力大幅提高我們網絡的整體資源利用率，直接降低向客戶交付每token的實際成本。

LLM的分佈式推理加速

模型優化。我們通過軟硬件創新相結合，增強了廣泛採用的開源模型。通過智能合併計算任務、簡化計算邏輯以及採用推測性採樣等技術，我們將模型輸出效率提高了7倍以上，並將理論運營成本降低85.7%。這有助於在幾乎不犧牲質量的情況下實現更快、更實惠的AI性能。

業 務

分佈式推理優化。我們的優化能力涵蓋算子優化、推理引擎優化、存儲系統優化、為前沿硬件量身定制的模型優化等多個維度。通過該等多層次的優化，我們能夠將計算資源利用率推至接近理論極限，從而大幅降低成本。例如，於2026年，通過我們的模型API服務交付的知名大模型的輸入代幣定價較主要國際雲服務商的平均收費低40%以上。

我們的優化能力涵蓋多個層面，例如算子優化、推理引擎優化、儲存系統優化及為尖端硬件定制的模型優化。通過該等多層優化，我們能夠將計算資源利用率推至接近其理論極限，大幅降低成本。

關鍵運營指標

我們使用以下關鍵運營指標管理業務。我們使用該等指標評估業務進展，並就如何分配資本、時間及技術投資作出決策。

	截至12月31日／截至該日止年度			截至2026年 4月30日／ 截至該日 止四個月
	2023年	2024年	2025年	
邊緣雲計算服務				
計算節點數量 ⁽¹⁾	2,849	4,012	4,632	4,956
我們的邊緣雲計算節點 覆蓋的縣市數量 ⁽²⁾	999	1,296	1,345	1,441
CPU核心數量 ⁽³⁾	299,166	387,011	446,525	479,148
服務器儲存總量(TB) ⁽³⁾	57,693	64,390	92,230	98,700
付費客戶數量	23	23	32	23
AI雲計算服務				
計算節點數量 ⁽¹⁾	5	19	64	77
我們的AI雲計算節點 覆蓋的縣市數量 ⁽²⁾	5	15	42	53
開源模型數量 ⁽⁴⁾	—	82	176	146
註冊開發者數量 ⁽⁵⁾	12,112	125,545	524,436	574,485
活躍用戶數量 ⁽⁶⁾	12,169	114,692	180,851	59,020
付費客戶數量 ⁽⁷⁾	7,454	47,860	91,538	13,832
總算力 (PFLOPS) ⁽⁸⁾	238	1,482	17,430	25,793
每名用戶平均購買額 ⁽⁹⁾	—	—	226.065	352.018

附註：

- (1) 由於我們按輕資產模式營運，絕大部分計算節點均為租賃。
- (2) 截至2025年12月31日，我們用於邊緣雲計算服務的所有計算節點均位於中國；而我們用於AI雲計算服務的計算節點中，26個位於中國、8個位於美國、2個位於英國，而在加拿大、新加坡、冰島、葡萄牙、印度尼西亞及馬來西亞各有1個。
- (3) 指相關年度／期間最後一個月的交易量。
- (4) 指於相關年度／期間，透過我們的模型API服務公開釋出的開源模型數量。由於我們移除了使用率低、重複及實驗性模型，以便將資源集中於核心旗艦模型，因此開源模型數量由2025年至截至2026年4月30日止四個月呈現下降趨勢。此舉並未影響平台使用率：同期token消耗量與總運算能力持續增長，顯示模型數量減少並未降低用戶需求。相反地，此項調整標誌著策略重心已從擴展模型數量，轉向優化模型質量、推論效率及商業轉化率。

業 務

- (5) 指分別截至2023年、2024年及2025年12月31日以及2026年4月30日已於我們平台註冊的開發者。
- (6) 指於2023年、2024年及2025年以及截至2026年4月30日止四個月使用我們模型API服務的用戶。截至2026年4月30日止四個月的活躍用戶數量並未按年化計算，因此較低。
- (7) 由2025年至截至2026年4月30日止四個月，我們的付費客戶數量減少，這主要與開源模型數量減少的趨勢一致。詳情請參閱上文附註(4)。因此，低付費客戶數量有所減少，但部分被高付費客戶數量的增加所抵銷。此客戶結構的轉變，反映了我們持續將商業重心放在提升營收質量及變現效率上。
- (8) PFLOPS為用於描述極大計算能力的標準單位。1 PFLOPS相當於每秒 10^{15} 次浮點運算。
- (9) AI雲計算服務的每名用戶平均購買額，乃按於特定年度／期間消費超過人民幣1,000元的用戶所產生的總收入，除以該年度／期間該等用戶的平均數目計算。

此外，就我們的AI雲計算服務分部而言，日均token消耗量（指我們的AI雲計算平台於某一期間每日處理的token平均數量）由2024年12月的271億增加至2025年12月的2,710億，並進一步增加至2026年4月的10,280億。

研發

我們濃厚的創新熱情加上強大的研發能力，使我們能夠跟上雲計算行業技術的快速發展。截至2023年、2024年及2025年12月31日止年度，我們產生的研發開支分別為人民幣67.8百萬元、人民幣86.3百萬元及人民幣119.9百萬元，分別佔我們同年收入的18.9%、15.5%及15.6%。

我們高度重視招聘在行業內具有豐富經驗的技術專家及高級工程師。我們的技術能力乃由我們才華橫溢、兢兢業業的研發團隊建立。截至2025年12月31日，我們擁有一支由170名研發專業人士組成的團隊，佔我們僱員總數約61.8%。我們已設立多項培訓計劃，以使我們的研發人員了解相關領域的最先進技術。於往績記錄期，我們主要實施自主研發。

研發流程

我們研發流程的主要階段包括：

- **概念化階段。**我們的研發團隊根據現有客戶及行業利益相關者的意見，確定市場需求及技術差距。我們就候選產品應達到的水平設立清晰的目標。
- **可行性研究。**我們會從技術及商業兩個角度考察候選產品。技術可行性涉及評估可用技術、與現有系統的集成度及兼容性，而經濟可行性則評估成本及資源需求。
- **設計與開發。**我們設計候選產品的架構，其中包括軟件、硬件及算法等組件。此外，我們開發一個原型，並對功能、安全性及可擴展性進行嚴格測試。該原型還需要反覆修改，以完善功能並解決問題。
- **部署測試。**我們會發佈試用版，供選定的用戶進行實際場景測試。我們會收集有關可用性的反饋並評估產品的性能。我們亦進行壓力測試以評估產品在高需求下的可靠性及可擴展性。
- **發佈。**我們向所有用戶推出功能齊全的產品。我們亦制定營銷策略以宣傳產品的獨特功能及優勢，並確保向客戶提供充分且及時的客戶支持。

業 務

重點研發項目

以下載列我們主要研發項目的例子：

- **人工智能客戶服務系統**。此項目專注於通過將大型語言模型與業務工作流程深度融合，開發超越腳本式回應的新一代人工智能客服平台。我們的目標是建立一個能夠自主決策並主動為客戶服務的系統，而不只是回應查詢。
- **CDN流量調度系統**。此項目旨在開發一種低成本、高性能且易於操作及維護的CDN產品。研究集中在三個方面，第一，根據網絡資源的特點對網絡資源進行分類，建立成本模型以優化帶寬分配。第二，利用歷史流量數據更有效地預測未來需求模式並指導帶寬規劃。第三，構建可視化、容易使用的運行管理系統，提高網絡變更的安全性、準確性及速度。
- **虛擬機器平台**。此項目將虛擬化技術應用於我們現有的物理計算資源，使其能夠劃分為多個虛擬機器。這為交付及運作服務創造更靈活的新方式，並提高整體資源利用率。

知識產權

我們的專利、版權、商標、商業秘密及其他知識產權對我們的業務營運至關重要。我們主要依靠專利、版權、商標、商業秘密、反不正當競爭法及合約權利（如與僱員及客戶訂立的保密協議）來保護我們的知識產權。我們在僱傭協議及商業協議中明確載列有關我們知識產權的所有權及保護的所有權利及義務。此外，我們已實施一套全面措施以保護我們的知識產權。

截至2025年12月31日，我們在中國擁有34項專利、37個商標、32個域名及67個版權。

於往績記錄期及截至最後實際可行日期，我們並無因侵犯第三方知識產權而產生任何重大爭議或索償。

數據隱私及數據安全

數據安全及保護是我們的首要任務。就此而言，我們已獲得ISO 27001信息安全管理体系認證。我們已制定嚴格的數據保護及信息安全政策，以確保嚴格遵守適用的法律、法規及現行行業慣例。

數據收集

我們在提供邊緣雲計算服務時，為整合第三方提供商的計算資源以滿足合作客戶對邊緣計算及傳輸以及完成交易的需求，我們一般收集的數據包括：(i)客戶的基本信息、付款信息及聯絡信息；(ii)算力供應商的基本信息、付款信息、聯絡信息、設備信息及資格證書。

我們在提供AI雲計算服務時，除收集上述數據外，亦會透過我們的網站收集若干個人信息，例如手機號碼、用戶姓名、電郵地址以及身份認證信息（例如真實姓名及身份證號碼），並詳細說明我們收集個人信息的情況。我們僅在進行賬戶註冊及登錄、完

業 務

成客戶（包括公司客戶及個人開發商用戶）實名身份驗證、向其提供相關服務及回應服務諮詢所需的範圍內處理該等個人信息。

於往績記錄期，我們並無向客戶或任何其他第三方購買任何個人信息或其他數據。我們的客戶保留其數據的全部所有權及控制權。為免生疑問，由於我們僅作為中介技術服務提供商經營業務，故在使用我們的分佈式雲計算服務時無法獲取客戶或其終端用戶的數據。

數據儲存及刪除

我們在中國內地收到的所有信息及數據均在中國內地儲存及保存。我們根據不同的安全級別區分數據安全儲存。我們將根據相關法律及法規在數據保留期屆滿後刪除數據。

數據共享與傳輸

為滿足客戶對邊緣計算及傳輸的需求，我們的邊緣雲計算服務不會將數據傳輸至中央雲數據中心，而是以邊緣傳輸技術傳輸至邊緣服務器，將算力下沉到互聯網「邊緣」（即靠近數據生成及使用的位置）處理。我們的境內外系統分別部署在獨立的雲服務器，在實體和虛擬層面上相互隔離，並無數據由國內傳輸至海外。此外，我們已採納旨在防止非法及／或未經授權跨境數據傳輸的內部規則及程序。該等措施確保我們的業務營運並不涉及任何未經授權及／或非法跨境數據傳輸。

根據我們的內部控制規則、私隱政策及我們與客戶之間的合約，除非根據法律規定或我們已獲得客戶許可及授權，否則我們不會將客戶的數據或個人信息向第三方分享、披露或傳輸。

內部控制措施

我們已實施有關保護數據私隱及安全的內部政策，目的是確保數據和信息安全，優化數據治理，保護我們客戶、業務合作夥伴、僱員及其他第三方的利益，並確保遵守所有適用的法律及法規。我們實施一套健全的內部認證及授權系統，以確保我們的機密及重要業務數據及商業秘密僅可在獲授權使用的情況下及由獲授權人員查閱。我們已建立與數據安全要求、國家標準及行業最佳實踐相關的信息體系，並擬持續在數據安全及私隱保護方面投入大量資金。我們的信息系統採用多層保障措施，包括內部及外部防火牆，以識別及保護我們免受安全攻擊。我們的主要業務系統已根據《信息安全等級保護管理辦法》取得《信息系統安全等級保護備案證明》。我們在中國內地收到的所有信息及數據均已在中國內地儲存及保存。我們的境內外系統分別部署在獨立的雲服務器，在實體和虛擬層面上相互隔離，之間並無數據交換。此外，我們已採納旨在防止非法及／或未經授權跨境數據傳輸的內部規則及程序。該等措施確保我們的業務營運並不涉及任何跨境數據傳輸。

業 務

遵守中國及我們的主要海外市場的數據隱私及網絡安全法律

於往績記錄期及直至最後實際可行日期，(i)我們並無接獲任何第三方以侵犯中國或其他司法權區任何適用法律及法規規定的任何第三方的數據及隱私保護權為由向我們提出的任何申索；(ii)我們並無接獲任何有關數據隱私或安全措施的投訴；(iii)我們已實施有關保護數據隱私及安全的內部政策，目的是確保數據及信息安全並確保遵守所有適用法律及法規；(iv)於往績記錄期，概無發生重大數據或個人信息外洩事件；(v)於往績記錄期，據我們所深知，我們概無面臨政府主管部門或第三方對我們提起的有關違反相關網絡安全、數據安全及個人信息保護法律或法規的待決或可能提起的調查、法律程序或行政處罰；及(vi)我們將持續密切關注數據安全的監管動態，並遵守最新的監管要求。有鑒於此，我們的中國法律顧問認為，於往績記錄期及直至最後實際可行日期，我們於所有重大方面均遵守目前有效且適用的數據隱私及安全法律法規。

根據美國數據法律法律顧問的意見，我們遵守美國隱私法律的責任主要限於加州消費者隱私法(「CCPA」)(經《加州隱私權法案》(「CPRA」)修訂)。網絡安全方面，雖然美國並無獨立的全國網絡安全法律，但監管指引及執法機構(特別是聯邦貿易委員會(「FTC」))規定企業實施並維持切合所處理信息性質的合理安全程序及實務。基於美國數據法律法律顧問的評估，1)我們的實務符合或優於美國法律要求的「合理安全」最低標準；2)我們採取與服務商及客戶的合約、技術保障措施，進一步支持合規，及3)我們的業務運營符合CCPA/CPRA以及同類美國隱私及網絡安全法律規定的服務商責任。再者，美國數據法律法律顧問認為，美國聯邦及州隱私及網絡安全法律並無規定可能對我們的業務運營造成重大不利影響的責任。

基於新加坡數據法律法律顧問的意見，2012年個人資料保護法(「個人資料保護法」)是新加坡有關數據隱私、安全及跨境傳輸的主要法例。個人資料保護法載列廣泛適用於在新加坡收集、使用或處理個人資料的組織的通用合規責任。於往績記錄期及直至新加坡數據法律法律顧問意見日期，我們未曾受到根據個人資料保護法施加的執法行動、調查、監管指令或經濟處罰。網絡安全方面，新加坡數據法律法律顧問告知，由於我們及我們的產品或服務並非2018年新加坡網絡安全法規定的指明關鍵信息基礎設施，且我們並無提供須獲許可的網絡安全服務，我們合理預計，其中規定的監管框架不會對我們的業務運營造成任何重大不利影響。

算法和人工智能合規

近年來，中國當局頒佈了關於中國算法和人工智能法規的若干規定，包括《互聯網信息服務深度合成管理規定》(「深度合成管理規定」)，自2023年1月10日起生效；《生成式人工智能服務管理暫行辦法》(「人工智能辦法」)，自2023年8月15日起生效；及《人工智能生成合成內容標識辦法》(「標識辦法」)，於2025年9月1日起生效。詳情請參閱本文件「監管概覽－有關生成式人工智能服務的法規」。

在我們提供的產品和服務中，只有模型API涉及生成式人工智能服務，該服務僅限以下用戶使用：(i)企業客戶；及(ii)有限試用的個人開發者用戶。我們提供模型API適用深度合成管理規定和標識辦法。根據人工智能辦法，確定人工智能辦法適用性有

業 務

兩個要素：(i)服務是否為生成式人工智能服務；及(ii)此類服務是否在中國境內向公眾提供。對此，尚未有官方解釋或具體實施規則；關於我們的模型API是否被視為向公眾提供服務以及因此受到人工智能辦法的約束，本質上仍未明確。

儘管存在不確定性，但截至最後實際可行日期，我們已採取一系列措施（考慮到人工智能辦法適用於我們的可能性），以遵守上述與算法和人工智能相關的法規，包括但不限於：(i)我們建立了專門的算法安全專職機構；(ii)我們對算法進行了定期的安全測試和評估；(iii)我們建立了信息安全監控機制來審查生成的內容，為投訴和報告非法信息提供了渠道，當我們發現非法內容時，我們將及時處理；(iv)我們制定並實施了《算法安全監控制度》、《算法安全自我評估制度》、《算法違法違規處置制度》、《算法安全事件應急處置制度》和《科技倫理審查制度》等內部措施；(v)我們已經在網信辦的互聯網信息服務算法備案系統中完成了三個深度合成服務算法的備案，並在我們網站顯著位置標明備案編號，並提供了公示信息鏈接；(vi)我們為人工智能生成的合成內容添加了顯式或隱式標籤，並實施了相關的技術準備；(vii)我們與中國相關監管部門保持溝通，以充分理解並遵守最新的合規要求。董事認為，即使人工智能辦法適用於我們，並且標識辦法正式施行，此類規定也不會對我們產生重大不利影響。基於上述情況，我們的中國法律顧問、聯席保薦人和董事認為，於往績記錄期及直至最後實際可行日期，我們在所有重大方面都遵守了上述與算法和人工智能相關的法規。

銷售及營銷

截至2025年12月31日，我們擁有一支由43名人員組成的銷售及營銷團隊。我們實施產品主導型增長及開發者友好型戰略以吸引客戶。我們通過參與開源社區活動及提供開源集成以吸引開發者使用我們的產品及服務。該策略有助我們以較低成本獲得新客戶，並在開發者社區內擴大我們的影響力。我們亦通過利用網絡效應及利益相關方在整個商業週期的推薦，並通過實地考察、行業會議及其他營銷活動進行銷售，從而戰略性地擴大我們的市場份額，並以具成本效益的方式擴大我們的業務規模。

定價

我們就特定產品或服務採用不同定價模式。下表載列我們各項主要產品或服務的定價模式：

產品／服務	定價模式
邊緣雲計算服務	
邊緣節點服務.....	按實際使用量收費，介於每MB人民幣2元至人民幣4元。
邊緣CDN	按實際使用量收費，介於每MB人民幣2元至人民幣4元。
AI雲計算服務	
GPU雲服務.....	按使用小時數收費（即總價格=GPU單價×GPU數量×使用時間）

業 務

產品／服務

定價模式

模型API	按實際消耗token、生成圖片張數或生成視頻時長收費。尤其是，(i)大型語言模型服務的費用根據每次互動中消耗的輸入token和輸出token的數量分開計算。總成本=(輸入token數目×輸入單位價格)+(輸出token數目×輸出單位價格)；及(ii)對於多模式模型，圖像生成按圖像計費；視頻生成以視頻為單位計費，並根據輸出的時長和分辨率遞增定價；及語音合成服務主要按每個字符收費。
智能體基礎設施	沙箱服務按秒計費，分別按虛擬CPU核心佔用率及內存容量獨立計算；沙箱實例終止後，計費將停止。

客戶支持及售後

我們深明，卓越的客戶支持是與客戶建立持久夥伴關係及逐步加強品牌聲譽的基石。我們快速響應及專業的售後服務不僅可提高客戶滿意度和參與度，而且對於持續保持強勁的客戶保留率方面發揮關鍵作用。我們致力於提供快速可靠的協助。我們的支持團隊快速響應客戶查詢並提供持續的進度更新。對於具有更複雜售後需求的客戶，我們的專業項目經理提供全面的全生命週期支持。他們提供量身定制的技術諮詢及實踐指導，使我們的服務能夠與客戶的獨特業務需求無縫整合。透過持續的溝通和周到的服務，我們深入了解客戶不斷變化的期望和業務挑戰。這種積極主動的溝通不僅使我們能夠預測未來的需求，更使我們能夠發掘新的商機，從而鞏固我們作為值得信賴、具有前瞻性的商業合作夥伴的地位。

我們的客戶

我們的客戶群體廣泛，涵蓋各行各業，包括泛娛樂、社交媒體、電商、教育、汽車、電信、信息科技、智能製造等領域。隨著我們產品線的擴充和服務區域的延伸，我們的客戶群持續擴大。我們在服務好大客戶的同時，積極儲備成長中的客戶，以促進業務持續增長。

於2023年、2024年及2025年，我們的收入分別為人民幣358.4百萬元、人民幣558.0百萬元及人民幣770.3百萬元。於2023年、2024年及2025年，我們來自最大客戶的收入分別佔我們同年收入的44.1%、35.2%及30.0%。於2023年、2024年及2025年，我們來自五大客戶的收入分別佔我們同年收入的92.5%、89.5%及79.0%。請參閱「風險因素－與我們的業務有關的風險－一名或多名主要客戶的流失或其使用量大幅減少將導致收入減少，並可能損害我們的業務」。

我們與最大客戶（客戶A及客戶B）的關係

截至2023年、2024年及2025年12月31日止年度，我們來自客戶A的收入分別佔我們總收入的44.1%、35.2%及30.0%；我們來自客戶B的收入分別佔我們總收入的27.7%、32.6%及23.3%。

業 務

董事認為，我們於往績記錄期的客戶集中主要歸因於我們與主要客戶建立並維持穩固且長期關係的能力。截至最後實際可行日期，我們與客戶A及客戶B的業務關係分別約為五年及六年。我們相信，通過提供一套全面的邊緣雲計算及AI雲計算服務，持續且繼續滿足他們不斷變化的技術要求、性能標準及成本預期，我們維持了主要客戶的忠誠度。此外，即使我們的主要客戶已擁有其自身的計算基礎設施，但彼等仍在需求高峰期間依賴我們補充其容量。我們提供深度技術整合、積極響應的客戶支持及持續服務優化的能力，使我們多年來能夠從該等客戶獲得經常性及不斷增長的商機。於往績記錄期，我們並無與於往績記錄期任何年度為我們五大客戶之一的任何客戶產生任何重大糾紛。

儘管於往績記錄期我們的收入集中於最大客戶，我們基於以下理由認為我們的業務可持續：

- (i) 除客戶A及客戶B外，於往績記錄期任何年度，我們五大客戶中概無任何客戶佔我們於往績記錄期該年度的總收入超過20%；
- (ii) 於整個往績記錄期，我們與五大客戶均維持穩定及長期的業務關係。截至最後實際可行日期，我們與客戶A及客戶B的業務關係分別約為五年及六年，而與該等客戶的長期合作已帶來深度的技術、服務及產品整合；
- (iii) 我們相信，維持該等業務關係對我們的主要客戶及我們均屬互惠互利。根據灼識諮詢的資料，中國的主要互聯網公司為向其客戶提供低延遲的內容傳輸，對邊緣雲計算服務有重大需求。主要互聯網公司一般傾向與具備成熟技術能力及擁有經驗證大規模基礎設施的雲服務提供商合作，因為該等供應商更能提供大型平台工作負載所需的性能、可靠性及成本效益，並能在不出現服務降級的情況下應對計算需求的快速波動。因此，根據灼識諮詢的資料，雲計算服務提供商的客戶集中於主要互聯網公司乃屬行業慣例。我們相信，我們專有的分佈式計算架構、領先的調度能力、分佈式推理加速技術、廣泛分佈於全球的計算節點網絡，以及提供大規模可靠且具成本效益服務的能力，構成了我們與主要客戶穩固關係的基石，而董事相信該等關係將繼續保持穩定；及
- (iv) 我們的客戶集中度於往績記錄期已有重大改善。我們五大客戶的收入貢獻由2023年的92.5%下降至2024年的89.5%，並進一步下降至2025年的79.0%，反映了我們在不同垂直行業的客戶群穩步擴大，以及我們AI雲計算服務的快速增長。

我們將通過持續技術創新、與客戶聯合開發、拓展至新的垂直行業及地域市場，以及進一步開發我們的AI雲計算服務，持續多元化並擴大我們的客戶群。特別是，我們的AI雲計算服務於往績記錄期實現了快速增長，吸引了來自不同領域且不斷擴大的客戶群，該等客戶與我們既有的邊緣雲計算客戶群有所區別且形成互補。此種橫跨兩個不同且相互促進業務分部的客戶畫像多元化，顯著降低了我們對任何單一客戶或垂直行業的依賴，且我們預期隨著AI在各行業及地域採用的深化，此趨勢將會持續。我們仍致力於不斷加強我們的技術能力，擴展全球計算網絡及提升我們的服務項目的深度，以吸引並留住不斷增長的高質量客戶群，從而減輕我們與主要客戶關係發生任何重大不利變動或終止的風險。

業 務

以下為往績記錄期各年來自五大客戶的收入明細及彼等各自的背景資料：

截至2023年12月31日止年度：

排名	客戶	提供服務的類型	交易金額 (人民幣千元)	佔總收入的 百分比 (%)	背景及主要 業務活動	與本集團關係 開始年份	信貸期
1	客戶A	邊緣雲計算服務	158,195.9	44.1	一家總部位於北京的中國公司，專注於一系列內容平台開發及運營	2021年	60天
2	客戶B	邊緣雲計算服務	99,348.2	27.7	一家總部位於北京的中國公司，專注於提供網絡技術服務、軟件開發及數據處理服務	2020年	30天
3	客戶C	邊緣雲計算服務	51,224.2	14.3	一家總部位於深圳的中國公司，專注於通信和社交媒體服務	2021年	28天
4	客戶D	邊緣雲計算服務	13,100.2	3.7	一家總部位於北京的中國公司，專注於提供互聯網相關服務、包括搜尋引擎服務及在線營銷	2021年	30天
5	客戶E	邊緣雲計算服務	9,765.9	2.7	一家總部位於長沙的中國公司，專注於文化和娛樂業務	2021年	60天
總計			<u>331,634.4</u>	<u>92.5</u>			

業 務

截至2024年12月31日止年度：

排名	客戶	提供服務的類型	交易金額 <i>(人民幣千元)</i>	佔總收入的 百分比 <i>(%)</i>	背景及主要 業務活動	與本集團關係 開始年份	信貸期
1	客戶A	邊緣雲計算服務	196,176.4	35.2	一家總部位於北京的中國公司，專注於一系列內容平台開發及運營	2021年	60天
2	客戶B	邊緣雲計算服務	182,169.8	32.6	一家總部位於北京的中國公司，專注於提供網絡技術服務、軟件開發及數據處理服務	2020年	30天
3	客戶C	邊緣雲計算服務	58,526.1	10.5	一家總部位於深圳的中國公司，專注於通信和社交媒體服務	2021年	30天
4	客戶D	邊緣雲計算服務	32,761.3	5.9	一家總部位於北京的中國公司，專注於提供互聯網相關服務、包括搜尋引擎服務及在線營銷	2021年	28天
5	客戶F	邊緣雲計算服務	29,440.8	5.3	一家總部位於上海的中國公司，專注於數字娛樂平台開發及運營	2020年	90天
總計			<u>499,074.4</u>	<u>89.5</u>			

業 務

截至2025年12月31日止年度：

排名	客戶	提供服務的類型	交易金額 (人民幣千元)	佔總收入的 百分比 (%)	背景及主要業務活動	與本集團關係 開始年份	信貸期
1	客戶A	邊緣雲計算服務	230,861.3	30.0	一家總部位於北京的中國公司，專注於一系列內容平台開發及運營	2021年	60天
2	客戶B	邊緣雲及AI雲計算服務	179,473.2	23.3	一家總部位於北京的中國公司，專注於提供網絡技術服務、軟件開發及數據處理服務	2020年	30天
3	客戶C	邊緣雲計算服務	106,752.8	13.9	一家總部位於深圳的中國公司，專注於通信和社交媒體服務	2021年	28天
4	客戶F	邊緣雲及AI雲計算服務	61,387.0	8.0	一家總部位於上海的中國公司，專注於數字娛樂平台開發及運營	2020年	90天
5	客戶D	邊緣雲計算服務	29,777.2	3.9	一家總部位於北京的中國公司，專注於提供互聯網相關服務、包括搜尋引擎服務及在線營銷	2021年	30天
總計			<u>608,251.5</u>	<u>79.0</u>			

我們於往績記錄期各年的五大客戶均為獨立第三方。於往績記錄期各年及截至最後實際可行日期，概無董事、彼等的緊密聯繫人或任何股東（就董事所知，擁有我們已發行股本5%以上的權益）於我們五大客戶中擁有任何權益。

於往績記錄期，我們與客戶訂立的標準協議的主要條款概要如下：

- **範圍。**我們的服務範圍一般載於服務條款內。

業 務

- **定價。**我們主要根據數據資源用量、數據流量及帶寬按實際用量向客戶收費。詳情請參閱「業務－銷售及營銷－定價」。
- **付款及信貸條款。**我們每個月向客戶寄發發票，詳細列明彼等上個月對我們服務的實際使用情況。客戶通常須按月結算發票。我們一般向客戶授出30至90天的信貸期。
- **保密性。**各方均有責任於合約期間及期後對由另一方向其披露的所有機密資料嚴格保密。
- **期限及終止。**協議一般為期一年。任何一方於提前30天發出書面通知或以代通知金形式終止協議。協議期限可按相同條款延長，直至我們的客戶停止使用服務為止。

於往績記錄期及直至最後實際可行日期，我們在所有重大方面均遵守與主要客戶的協議條款，且我們並無遭遇任何導致與主要客戶發生任何合約糾紛或主要客戶提出索賠的情況而對我們的經營造成重大不利影響。

我們的供應商

我們的供應商主要為算力資源的供應商。於2023年、2024年及2025年，我們與最大供應商的交易金額分別佔我們同年銷售成本的4.0%、6.0%及5.5%。於2023年、2024年及2025年，我們與五大供應商的交易金額分別佔我們銷售成本的15.2%、20.3%及19.5%。於往績記錄期，我們一般以銀行轉賬方式向供應商結算付款。於往績記錄期各年，我們與五大供應商建立並保持穩定及良好的關係，與其中大部分供應商擁有四年或以上的合作關係。我們的主要供應商位於中國，詳情請參閱下文。

以下為我們五大供應商於往績記錄期各年的明細及彼等各自的背景資料：

截至2023年12月31日止年度：

排名	供應商	提供服務的類型	佔銷售成本的		背景及主要 業務活動	與本集團關係 開始年份	信貸期
			交易金額	百分比			
			(人民幣千元)	(%)			
1...	供應商A	算力資源	11,869.2	4.0	一家總部位於北京的中國公司，專注於技術服務、數據處理	2020年	15天
2...	供應商B	算力資源	11,380.4	3.9	一家總部位於深圳的中國公司，專注於信息技術諮詢服務	2022年	15天

業 務

排名	供應商	提供服務的類型	佔銷售成本的		背景及主要 業務活動	與本集團關係 開始年份	信貸期
			交易金額	百分比			
			(人民幣千元)	(%)			
3	供應商C	算力資源	10,787.0	3.7	一家總部位於安陽的中國公司，專注於軟件開發、大數據服務	2021年	15天
4	供應商D	算力資源	5,545.0	1.9	一家總部位於廣州的中國公司，專注於IDC服務和信息技術	2022年	15天
5	供應商E	算力資源	4,896.2	1.7	一家總部位於南京的中國公司，專注於計算機網絡維護服務、技術服務	2021年	15天
總計			<u>44,477.8</u>	<u>15.2</u>			

截至2024年12月31日止年度：

排名	供應商	提供服務的類型	佔銷售成本的		背景及主要 業務活動	與本集團關係 開始年份	信貸期
			交易金額	百分比			
			(人民幣千元)	(%)			
1	供應商F	算力資源	29,375.6	6.0	一家總部位於保定的中國公司，專注於互聯網數據服務	2021年	15天
2	供應商A	算力資源	23,128.1	4.8	一家總部位於北京的中國公司，專注於技術服務、數據處理	2020年	15天
3	供應商G	算力資源	18,636.3	3.8	一家總部位於杭州的中國公司，專注於軟件和信息技術服務	2022年	15天
4	供應商H	算力資源	14,660.0	3.0	一家總部位於北京的中國公司，專注於邊緣雲服務、CDN加速、網絡安全和大數據	2023年	15天

業 務

排名	供應商	提供服務的類型	交易金額 (人民幣千元)	佔銷售成本的 百分比 (%)	背景及主要 業務活動	與本集團關係 開始年份	信貸期
5	供應商D	算力資源	13,405.2	2.7	一家總部位於廣州的中國公司，專注於IDC服務和信息技術	2022年	15天
總計			<u>99,205.2</u>	<u>20.3</u>			

截至2025年12月31日止年度：

排名	供應商	提供服務的類型	交易金額 (人民幣千元)	佔銷售成本 的百分比 (%)	背景及主要業務活動	與本集團關係 開始年份	信貸期
1	供應商I	算力資源	38,502.7	5.5	一家總部位於廣州的中國公司，專注於軟件開發服務	2025年	30天
2	供應商J	算力資源	32,505.3	4.7	一家總部位於廣州的中國公司，專注於信息技術諮詢服務	2025年	30天
3	供應商K	算力資源	23,608.6	3.4	一家總部位於山東的中國公司，專注於互聯網信息服務	2023年	15天
4	供應商F	算力資源	22,741.0	3.3	一家總部位於保定的中國公司，專注於互聯網數據服務	2021年	30天
5	供應商G	算力資源	18,333.2	2.6	一家總部位於杭州的中國公司，專注於軟件和信息技術服務	2022年	30天
總計			<u>135,690.8</u>	<u>19.5</u>			

業 務

我們於往績記錄期各年的五大供應商均為獨立第三方。於往績記錄期各年度及截至最後實際可行日期，概無董事、彼等的緊密聯繫人或任何股東（就董事所知，擁有我們已發行股本5%以上的權益）於我們五大供應商中擁有任何權益。

於往績記錄期，我們與主要供應商訂立的協議的主要條款概要載列如下：

- **範圍。**我們的採購範圍載於服務條款內。
- **定價。**我們通常按實際使用量收費。
- **付款及信貸條款。**我們一般須於供應商就上個月應計的服務開具發票後按月向供應商結算付款。
- **保密。**各方均有責任於合約期間及期後對另一方知悉的所有機密資料嚴格保密。
- **期限及終止。**協議一般為期一年。任何一方均可在雙方同意後終止協議。除非雙方另行協定，協議的期限通常可按相同條款延長一年。

於往績記錄期及直至最後實際可行日期，我們在所有重大方面遵守與主要供應商的協議條款，且我們並無遭遇任何導致與主要供應商發生任何合約糾紛或由主要供應商提出索賠的情況而對我們的營運造成重大不利影響。

重疊客戶及供應商

於往績記錄期，(i)客戶A為我們2023年、2024年及2025年的五大客戶之一，亦為我們2025年的供應商；(ii)客戶C為我們2023年、2024年及2025年的五大客戶之一，亦為我們同年的供應商；(iii)客戶D為我們2023年、2024年及2025年的五大客戶之一，亦為我們2025年的供應商；(iv)供應商E為我們2023年的五大供應商之一，亦為我們2023年的客戶；及(v)供應商H為我們2023年的五大供應商之一，亦為我們2023年、2024年及2025年的客戶（統稱為「重疊客戶－供應商」）。我們主要採購算力資源，並提供邊緣雲計算服務予重疊客戶－供應商。該等重疊客戶－供應商在非高峰時期擁有過剩的算力資源，但在高峰時期卻需要額外的算力資源。因此，在需要時我們既向其採購算力資源，亦向其供應算力資源。根據灼識諮詢的資料，該模式符合目前行業慣例。下表載列我們於往績記錄期向重疊客戶－供應商的銷售及採購情況：

	截至12月31日止年度		
	2023年	2024年	2025年
向重疊客戶－供應商銷售			
收入（人民幣千元）.....	59,697.2	64,350.3	305,421.4
佔總收入的百分比(%)	16.7	11.5	39.6
向重疊客戶－供應商採購			
採購金額（人民幣千元）.....	15,010.0	25,453.8	43,468.4
佔銷售成本的百分比(%)	5.1	5.2	6.2

業 務

董事確認，我們向重疊客戶－供應商的銷售及採購(i)在計及有關時間的通行採購及銷售價格的情況下經適當考慮後訂立，及(ii)按公平基準進行。於往績記錄期，重疊客戶－供應商所收取／支付的價格與其他客戶及供應商所支付／收取的價格相若。據董事所深知，於往績記錄期及直至最後實際可行日期，我們的其他主要客戶與主要供應商之間並無任何其他重疊。

競爭

中國及全球分佈式雲計算服務行業競爭激烈。我們主要在產品功能及範圍、性能、服務的可擴展性及可靠性、技術優勢、營銷及銷售能力、用戶體驗、定價、品牌知名度及聲譽方面進行競爭。此外，新興及增強技術可能進一步加劇行業競爭。有關詳情，請參閱「行業概覽－中國邊緣雲計算服務的競爭格局」及「行業概覽－中國獨立AI雲計算服務提供商的競爭格局」。

僱員

截至2025年12月31日，我們擁有275名全職僱員。下表載列截至2025年12月31日我們按職能劃分的僱員人數：

職能	截至2025年12月31日	
	數目	%
研發	170	61.8
銷售及營銷.....	44	16.0
營運	17	6.2
行政及其他.....	44	16.0
總計	275	100.0

我們相信，我們的成功取決於我們吸引、挽留及激勵合格人才的能力。我們主要透過招聘機構、校園招聘會、內部推薦計劃及線上渠道(包括本公司網站及社交網絡平台)招聘僱員。作為我們招聘及挽留策略的一部分，我們已制定全面的培訓計劃，涵蓋企業文化、僱員權利及責任、團隊建設、合規及工作表現等主題。

我們提供具競爭力的薪酬、績效晉升、激勵性股票期權、花紅及其他獎勵。我們相信，於往績記錄期及直至最後實際可行日期，我們與僱員維持良好的工作關係，且我們並無遇到任何重大勞資糾紛，亦無在招募營運員工方面遇到任何困難。

我們與僱員訂立標準僱傭、保密及不競爭協議。我們參與由適當的地方市級及省級政府設立的住房公積金及各類僱員社會保障計劃，包括住房、養老金、醫療、工傷及失業福利計劃。

業 務

物業

我們的總部位於中國上海。截至最後實際可行日期，我們並無擁有任何物業。截至2025年12月31日，我們於中國租賃六項物業，總建築面積約為2,912.44平方米。我們的租賃物業主要用作辦公室。

截至最後實際可行日期，根據適用的中國法律法規，我們在中國的租賃物業須向中國有關土地及房地產管理局辦理登記及備案，全部尚未辦理登記或備案。

據我們的中國法律顧問告知，未辦理租賃協議登記及備案不會影響租賃協議的有效性或導致我們須騰出租賃物業。然而，有關中國主管部門可就每項未登記租賃處以人民幣1,000元至人民幣10,000元不等的罰款。

經考慮上文所述，董事認為，上述租賃未登記事項不會個別或共同對我們的業務及經營業績造成重大影響，理由是：(i)於往績記錄期及直至最後實際可行日期，我們並無因未辦理相關租賃協議登記及備案而受到處罰，(ii)據我們的中國法律顧問告知，倘租賃登記可根據相關法律法規在主管政府部門責令的規定期限內完成，則政府部門就該等租賃物業對我們施加重大處罰的風險甚微。

保險

根據中國法律及法規的規定，我們的員工保險包括養老保險、失業保險、工傷保險及醫療保險。與一般市場慣例一致，我們並無購買任何營業中斷險或產品責任險，而該等保險根據中國法律並非強制性。於往績記錄期，我們並無就我們的業務提出任何重大保險索償。

我們相信我們的保險政策符合中國的相關規則及規例。有關詳情，請參閱本文件「風險因素－與我們的業務及行業有關的風險－我們可能並無足夠的保險範圍來承保潛在負債或損失，且倘產生任何該等負債或損失，我們的業務、財務狀況、經營業績及前景可能會受到重大不利影響。」一段。

牌照、許可證及批文

於往績記錄期及直至最後實際可行日期，我們已自相關監管機構取得對我們於中國的經營而言屬重要的所有必要牌照、批文及許可證。截至最後實際可行日期，下表載列我們的重要牌照及許可證：

牌照／許可證	持有人	頒發機構	頒發日期	到期日
增值電信業務經營許可證 (IDC/CDN/ISP)	派歐上海	中華人民共和國工業 和信息化部	2025年7月24日	2030年5月30日

業 務

牌照／許可證	持有人	頒發機構	頒發日期	到期日
增值電信業務經營許可證 (ICP).....	派歐上海	上海通信管理局	2025年5月26日	2030年7月17日
增值電信業務經營許可證 (ICP).....	生成時代	上海通信管理局	2025年6月30日	2030年6月30日

法律程序及不合規

於往績記錄期及直至最後實際可行日期，我們並無牽涉任何我們認為會對我們的業務、經營業績或財務狀況產生重大不利影響的實際或待決法律、仲裁或行政程序（包括任何行政處罰、破產或破產接管程序）。截至最後實際可行日期，我們並不知悉有任何我們認為會對我們的業務、經營業績或財務狀況產生重大不利影響的針對我們或任何董事的未決或具有威脅的法律、仲裁或行政程序。

於往績記錄期及直至最後實際可行日期，我們並無牽涉我們認為會對我們的業務、經營業績或財務狀況產生重大不利影響的任何重大不合規事件。

於往績記錄期及截至最後實際可行日期，我們的服務及解決方案概無出現任何重大錯誤、缺陷、安全漏洞或服務中斷，亦無任何客戶對我們提出任何重大申索。

違反第32號通知

於往績記錄期，我們從異構資源（包括並無持有電信服務牌照的實體）獲得未充分利用的計算資源（「該等事件」）。我們的分佈式計算技術使我們能夠有效地分配分散且利用不足的資源，並為客戶提供具成本效益的解決方案。鑒於涉及的供應商數量眾多，我們並未逐一核實每家供應商是否持有相關牌照。於2023年、2024年及截至2025年9月30日止九個月，我們分別委聘2,347家、3,723家及2,314家無牌供應商。於2025年，無牌供應商的數目大幅減少，因為我們於該期間開始終止與該等供應商的關係，而終止程序已於2025年前完成。於2023年、2024年及2025年，無牌供應商應佔的計算資源採購成本分別為人民幣151.4百萬元、人民幣189.1百萬元及人民幣154.1百萬元，佔我們相關年度總採購成本的54.3%、42.2%及28.5%。

誠如我們的中國法律顧問告知，(i)根據《工業和信息化部關於清理規範互聯網網絡接入服務市場的通知[2017]32號》（「第32號通知」），（其中包括）IDC、ISP及CDN企業不應使用無電信服務牌照的實體提供的帶寬資源。IDC、ISP及CDN企業須定期進行自查，並在必要時整頓其運營。我們在該等事件中使用帶寬資源並不構成中國法律及法規下的重大違規行為。根據第32號通知，該等事件並非歸入為「非法經營」，而是「須要糾正的行為」。我們的業務亦不構成第32號通知項下的「層層轉租」；(ii)於往績記錄期及直至最後實際可行日期，我們並無受到中國任何主管機關的任何行政處罰；(iii)根據與中國主管機關於2025年12月進行的電話訪談，主管機關確認我們不會因該等過

業 務

往事件而受到任何處罰，前提是我們已根據相關法律法規通過終止與所有無牌供應商合作的方式完成糾正；(iv)我們已取得經營邊緣雲計算服務業務所需的所有必要電信服務牌照；及(v)我們與我們的主要客戶於往績記錄期簽立的合約仍然具有法律約束力及可強制執行。

我們已相應實施糾正措施。截至2025年9月30日，我們已終止向全部不具備該等牌照的供應商進行採購。此外，我們已採用更嚴格的供應商資格審查標準。上海派雲為負責管理個別供應商的實體。我們已終止向全部此等無牌個別供應商採購，並將管理該等持牌供應商的責任轉交予派歐上海。上海派雲的唯一股東決議於2025年6月註銷上海派雲，且截至最後實際可行日期，有關註銷已根據相關法律法規完成。

基於我們目前擁有多元化的算力供應商團體，且中國存在大量未充分利用的計算資源，董事認為，該等事件及該等供應商的終止並未且不會對我們的業務及經營業績造成重大不利影響。我們來自邊緣雲計算服務的收入由2024年的人民幣547.7百萬元增加18.9%至2025年的人民幣651.1百萬元，並於2025年12月31日後繼續穩定運營。

為防止未來再次發生類似事件，我們已實施以下強化內部控制措施：

- 我們已採納算力採購管理政策，當中載列有關供應商准入及資格審查的明確及嚴謹要求。根據該政策，我們的採購部門必須在准入階段對潛在的算力供應商進行嚴格的盡職調查，包括收集及審查其營業執照、算力服務資格、技術能力認證及其他相關資格文件。
- 潛在供應商通過初步資格篩選後，將通過市場研究、招標或查詢流程進一步獲列入候選名單。我們的評估涵蓋價格競爭力（權重為40%）、算力表現、網絡質素及配套設施（合計權重為30%）以及服務回應速度與運維能力（權重為30%）。作為評估其中一環，我們的採購人員亦對獲列入候選名單的供應商的數據中心進行實地視察。
- 潛在算力供應商之後須接受我們技術部門及運維部門的聯合技術測試及評估。僅通過該聯合評估的供應商方可進入合約磋商階段。
- 於合約磋商階段，我們的採購部門磋商並界定服務範圍、售後支援安排及有關違約的合約補救措施。框架合約草擬稿將在簽立前由我們的法務部門審閱。
- 經我們的採購部門及技術部門批准後，我們的採購部門將在我們的供應商管理系統登記獲批准的算力供應商，以供日後採購之用。

業 務

環境、社會及企業管治

ESG治理

我們視ESG事宜為營運不可或缺的部分。我們致力於成為一家對社會負責的公司，不斷提升我們的ESG管理實踐，促進我們的長期可持續發展。

為加強對ESG事宜的管理能力，我們已建立ESG管理體系，由董事會統一領導、監督及實施。董事會確保我們的ESG策略與我們的整體業務目標一致，並負責審查及批准我們的ESG管理政策及ESG報告。

此外，我們已成立直接隸屬董事會的ESG工作小組，作為負責日常處理及協調ESG相關事宜的主要機構。ESG工作組負責了解利益相關方的訴求、意見及建議，協調內外部的相關工作，研究重要的ESG議題，指導ESG計劃的日常實施，並監督ESG報告的編製。各職能部門及附屬公司作為ESG工作的執行單位。各單位根據我們的整體計劃，負責落實ESG工作，並定期匯報工作進展。

ESG風險識別及評估

我們將ESG風險融入到企業戰略的頂層設計中，建立清晰的ESG管理方針及戰略。董事會發揮監督作用，定期檢討ESG事宜，以確保其符合我們的業務目標，有效控制風險及保障執行結果。

考慮到我們的業務特點和利益相關方的反饋，我們已建立全面的系統，用於識別並評估重大ESG議題。我們採用涵蓋影響程度、發生的可能性、時間敏感性等指標的多維度評分方法，有系統地評估相關風險及機遇。董事會定期領導程序，以確保風險優先級穩健及我們的應對策略針對性強，從而為我們的ESG風險管理提供持續優化的行動指引。通過該機制，我們能夠及時識別優先風險並制定相應的應對措施。

環境

我們基於分佈式計算技術建構全球算力調度平台。透過整合閒置計算資源並優化供需分配，我們有助於緩解全球計算能力結構性失衡的問題。借助該平台，我們將異構算力標準化為涵蓋雲端計算到人工智能推理的綜合服務組合，以滿足客戶多樣化的需求。

與此同時，我們亦意識到氣候變化對生態系統及商業的深遠影響。因此，我們已將氣候行動納入核心策略，致力於透過技術創新減少碳足跡，並將我們的算力網絡打造成數字化未來的綠色永續基石基礎設施。

業 務

氣候風險管理

我們意識到自身業務與氣候變化之間的密切聯繫，並已識別並評估可能影響我們業務運營、供應鏈及利益相關方的氣候相關風險。該等風險包括：

風險類別	潛在影響	管理措施
實體風險... 急性氣候風險，例如颱風及洪水以及長期風險，例如海平面上升及氣候模式的長期變化。	員工通勤受阻、辦公基礎設施受損、電腦設施及服務中斷，對業務連續性及業務運營造成不利影響。	我們持續完善緊急應變管理政策，優化緊急計劃及復修措施，以應對極端天氣事件，保障員工健康及安全，並確保客戶使用我們產品及服務的穩定性。
過渡風險... 更嚴格的碳排放法規帶來的政策風險，以及利益相關方更傾向於綠色發展帶來的聲譽風險。	潛在的運營挑戰及更高昂的合規成本，以及使我們的產品及服務符合低碳標準的壓力。	我們持續關注監管環境的變化，積極與利益相關方互動，並開發及實施更永續的生產及營運模式，以滿足日益變化的監管及市場期望。

我們透過設定環境目標來指導環境管理工作。我們已建立ESG目標審查機制，要求管理層定期檢視我們的目標。參考國內外ESG標準、產業同業以及我們的自身發展規劃，我們已制定以下環境目標：

- 溫室氣體減排目標：直至2030年每單位收入範圍1及範圍2的溫室氣體排放量比2024年減少10個百分比。
- 能源效益目標：直至2030年每單位收入的能源消耗量比2024年減少10個百分比。
- 用水效益目標：直至2030年每單位收入的耗水量比2024年減少10個百分比。
- 廢棄物減量目標：持續執行綠色辦公常規及減少產生廢棄物。

業 務

環境及氣候相關績效指標

我們將辦公室場所及算力中心納入氣候及環境績效指標內。鑒於我們業務的性質，我們並無從事實物產品製造，亦無經營我們自己的車隊。因此，我們不會產生空氣污染物或消耗包裝材料。

指標	單位	截至2023年 12月31日止年度	截至2024年 12月31日止年度	截至2025年 12月31日止年度
溫室氣體排放總量 ¹	噸二氧化碳當量	120.78	132.94	183.25
— 範圍2溫室氣體排放量 . .	噸二氧化碳當量	120.11	131.99	182.02
— 範圍3溫室氣體排放量 ² . .	噸二氧化碳當量	0.67	0.95	1.23
每單位收入的溫室 氣體排放量	噸二氧化碳當量／ 人民幣10,000元	0.34	0.24	0.24
間接能源消耗 ³	千瓦時	226,365.30	248,749.40	343,047.56
能源消耗總量	千瓦時	226,365.30	248,749.40	343,047.56
每單位收入的 能源消耗總量	千瓦時／ 人民幣10,000元	631.62	445.76	445.30
總廢水排放量	噸	740.0	1,055.0	1,365.00
每單位收入的總 廢水排放量	噸／ 人民幣1,000,000元	2.06	1.89	1.77
城市用水總量	噸	925.00	1,318.75	1,706.25
每單位收入的城市 用水總量	噸／ 人民幣1,000,000元	2.58	2.36	2.21

附註：

- (1) 我們並無擁有任何車輛，因此範圍1溫室氣體排放量不適用於我們。在計算我們與電力相關的溫室氣體排放量時，我們採用中華人民共和國生態環境部在《關於發佈2023年電力二氧化碳排放因子的公告》中公佈的電網排放因子。電力二氧化碳排放因子為0.5306 kg CO₂/MWh。
- (2) 我們的範圍3溫室氣體排放量的計算範圍包括處理我們排放的廢水所產生的溫室氣體排放。
- (3) 我們並無擁有任何車輛，我們亦無從事任何實物產品製造活動。我們的能源消耗僅包括我們辦公場所（包括照明、空調及辦公室設備）所用的電力。

我們的業務不涉及實體製造，因此我們對環境的影響相對有限。儘管如此，我們仍致力於綠色發展，嚴格遵守《中華人民共和國環境保護法》等環境法律法規。我們持續監察合規情況，並確保我們的營運符合適用規定。此外，我們已制定專門的環境保護政策，以系統地管理我們辦公室的環境事務。我們嚴格遵守《中華人民共和國固體廢物污染環境防治法》等廢棄物管理法律法規，並確保廢棄物的合規處理及處置。鑒於我們業務的性質，我們營運中產生的廢棄物主要包括生活廢棄物及辦公室廢棄物。所有無害廢棄物均由我們辦公樓宇的物業管理公司遵照適用要求收集及處理，而所有有害廢棄物則由我們委聘的合資格第三方服務供應商按適用要求處理。

業 務

社會

僱傭合規

我們制定並持續完善員工手冊等人事管理制度，以規範招聘及聘用的全過程。在招聘活動中，我們遵循公平、公正、公開、合適及擇優錄用的原則。我們已建立以工作要求為導向，並以工作能力及工作契合度為重點的評估標準。我們不會基於性別、年齡、種族、宗教或其他個人特徵對候選人進行區分，我們努力保障平等機會及多元化。

職業健康及安全

我們關心僱員的長遠利益，致力於構建完善的健康與安全保障體系。我們為所有僱員投保基本醫療保險及補充商業醫療保險，我們定期組織全體僱員健康檢查，以及早發現潛在的健康問題。我們亦為辦公室配備急救箱和基本急救藥品，並建立突發安全事件應急機制，確保僱員在需要時能及時得到現場幫助。截至報告期末，我們並無經歷任何與工作有關的受傷或死亡事件。

薪酬及福利

我們建立了公平合理且具有市場競爭力的薪酬管理體系，確保薪酬與工作職責、個人表現及貢獻相匹配。我們嚴格遵守國家對最低工資標準、工時制度及社會保險的規定，並為僱員按時足額繳納所有強制性社會保險及住房公積金，使其依法獲得勞動報酬及相關的社會保障權益。

我們持續完善僱員關懷機制，通過多樣化的福利待遇及文化活動，支持僱員在職業發展中平衡工作與生活。我們透過舉辦團隊建設活動、健康保障計劃及便利日常生活的措施，增強僱員的歸屬感及團隊凝聚力，同時培育積極、有復原力及具凝聚力的組織文化。

僱員發展及培訓

我們視人才發展為我們可持續競爭力的重要基礎。我們十分重視人才儲備及支持僱員的長期成長，致力於構建僱員與公司共同成長的組織生態系統。為此，我們已建立分級分類、覆蓋全職業週期、以實踐為導向的人才發展體系，並建立有系統的能力發展跟蹤機制。我們持續關注僱員專業技能及綜合能力的提升，以確保人才發展與我們的戰略發展方向保持一致。

截至2023年、2024年及2025年12月31日止年度，我們的全職員工培訓總人次分別為1,718人次、840人次及1,125人次，該等年度員工接受培訓佔比為62%、80%及69%。

供應鏈管理

我們視供應商為重要的戰略合作夥伴，致力於構建長期、穩定、協作及共贏的供應鏈生態系統。我們制定了《供應商管理實施細則》、《採購管理規定》等內部制度，以系統地規範我們的採購及供應商管理活動，落實我們的合規管理要求。

業 務

我們已建立以董事長辦公室為主導，跨職能協作為支撐的採購治理結構。該結構將採購管理委員會、審核小組以及資產管理、業務管理、質量管理、需求滿足等重點部門橫向聯繫起來，縱向明確不同層級在決策、執行、監督等方面的職責分工。由此形成職責清晰、協作高效的管理閉環。在供應商管理方面，我們實施分級分類管控，根據供應商風險等級及戰略重要性進行精細化管理，確保管控措施精準有效，從源頭提升供應鏈合規質量保證能力。

產品責任

我們嚴格遵守《中華人民共和國產品質量法》等適用的國家法律法規，在研發及營銷活動中始終堅持合規經營和審慎原則。我們將商業道德及合規經營作為企業發展的核心內容，在各項業務活動中嚴格遵守法律、法規及行業規範，從而不斷夯實產品責任管理基礎。為有系統地落實國家要求及行業標準，我們制定並實施研發項目管理辦法，以規範產品研發的整個過程，確保從需求分析及設計開發到測試及驗證的每個階段均可制度化、規範化的管理，為產品質量及市場合規提供堅實的保障。

反腐敗

我們視商業道德及合規經營為企業發展之根本。我們通過系統的內控流程和技術手段規範商業行為，堅決反對賄賂、欺詐、洗錢等違反商業道德和法律法規的行為。我們嚴格遵守《中華人民共和國刑法》、《中華人民共和國反洗錢法》等法律法規，並制定及實施《反欺詐、反洗錢和反賄賂管理制度》等內部政策。該等政策適用於全體員工，我們對董事、監事、中高級管理層以及關鍵崗位員工提出更高的合規要求，從而保障公司的持續、穩定、健康發展。

我們已建立以董事會直接為主導的內部審核體系，對商業道德與合規風險的防範、監督和處置進行監督。該體系運行獨立且具有權威性，並通過定期審核、專項檢查和流程控制等方式，有系統地識別潛在風險。對於重大或複雜事項，我們會聘請外部專業機構參與調查，以提高客觀性及可信度。

我們每年進行系統的風險評估。我們結合業務發展動態，識別潛在風險點，動態審視現有管控機制的有效性，並針對流程薄弱環節制定改進措施。同時，我們建立結合調查、問責、整改、糾正於一體的違規處置機制，確保風險的快速反應和有效控制。在責任追究上，我們區分領導責任和直接責任，對違規個人和相關管理人員按適用規定承擔責任，從而鼓勵部門加強日常管控。

業 務

風險管理及內部控制

我們致力構建並完善與業務營運相匹配的健全風險管理及內部控制體系，始終注重提高運作效能。我們持續檢討風險管理及內部控制政策及程序的執行情況，以提升其有效性及充分性。

財務報告風險管理

我們的財務報告風險管理包括一整套會計政策。我們已建立有效執行該等政策的程序，而我們的財務部門基於該等程序定期審閱管理賬目。此外，我們為財務部門員工提供持續培訓，以確保相關人員熟練掌握並能夠在日常營運中準確運用財務管理及會計政策。

內部控制風險管理

為確保遵守適用法規及內部標準，我們已制定嚴格的內部程序。我們的合規團隊與財務及業務部門緊密合作，以：(a)進行風險評估及就風險管理策略提供意見；(b)提高業務流程的效率及監察內部控制的成效；及(c)提高全公司風險意識。我們健全嚴格的內部程序以取得經營所需的所有牌照、許可證及批准，我們的內部控制團隊定期檢討以監察該等授權的狀態及有效性。我們的合規團隊亦與相關部門協調，以取得在有關部門備案所需的政府批准或同意。

人力資源風險管理

我們根據部門的不同需求開展常態化專項培訓，確保員工的技能水平與時俱進並符合客戶的服務目標。我們為員工提供員工手冊，員工手冊概述了有關最佳商業慣例、職業道德、防止欺詐、過失及貪污的內部規則及指引。此外，我們制定了商業行為及道德守則以及反賄賂和反貪污政策。該等指引概述最佳商業慣例及職業道德，並提供明確的反賄賂指引及措施。我們設置開放的內部舉報渠道，員工可舉報任何不法行為或不當行為，確保對所有舉報事件和個人進行調查，並根據調查結果採取適當行動。

獎項及認可

於往績記錄期，我們獲得有關服務、技術及創新的獎項及認可。下表載列我們獲得的部分主要獎項及認可的詳情：

獎項／認可	獲獎年份	頒獎機構／單位
極客傳媒「2025年度AI基礎設施卓越獎」.....	2025年	極客傳媒
量子位「2025人工智能年度傑出產品」.....	2025年	量子位
量子位「2025人工智能年度潛力創業公司」.....	2025年	量子位

業 務

獎項／認可	獲獎年份	頒獎機構／單位
三十六氦「WISE 2025商業之王年度最具商業潛力企業」.....	2025年	三十六氦
量子位「2025年值得關注的AIGC企業」.....	2025年	量子位
量子位「2024人工智能年度潛力創業公司」.....	2024年	量子位
中關村智用人工智能研究院「2024人工智能應用標桿」.....	2024年	中關村智用人工智能研究院
「2024中國科技產業引領榜」及《中國AI算力行業發展報告》重點廠商案例.....	2024年	甲子光年
SegmentFault「2024中國新銳技術先鋒企業榜單」	2024年	SegmentFault
中國信通院2024年度分佈式算力「星耀」案例	2024年	中國信通院
中國信通院2023年度算網基礎設施優秀案例 ...	2024年	中國信通院
IDC中國邊緣雲市場報告(前十)	2024年、2023年	IDC
中國信通院邊緣計算先鋒企業.....	2023年	中國信通院