

香港交易及結算所有限公司及香港聯合交易所有限公司對本公告的內容概不負責，對其準確性或完整性亦不發表任何聲明，並明確表示，概不就因本公告全部或任何部分內容而產生或因倚賴該等內容而引致的任何損失承擔任何責任。



Z.AI Co., Ltd.

北京智譜華章科技股份有限公司

(於中華人民共和國註冊成立之股份有限公司)

(股份代號：2513)

**截至2026年6月30日止六個月的
中期業績公告**

董事會欣然公佈本集團截至2026年6月30日止六個月的未經審計簡明綜合中期業績，連同2025年同期的比較數字。

本公告所載若干金額及百分比數字已作約整，或已約整至一位或兩位小數。任何表格、圖表或其他資料中所列總額與其內所列金額加總之間的任何差異，皆因約整所致。

財務摘要

	截至6月30日止六個月		同 比 變 動 %
	2026年	2025年	
	人 民 幣 千 元		
	(未 經 審 計)	(未 經 審 計)	
收 入	953,892	190,877	399.7
毛 利	251,611	95,424	163.7
研 發 開 支	(2,131,171)	(1,594,661)	33.6
期 內 虧 損	(2,071,992)	(2,357,852)	-12.1
經 調 整 期 內 淨 虧 損	(1,964,138)	(1,751,974)	12.1

	截至6月30日止六個月				收入增長的 同比變動 %
	2026年		2025年		
	人民幣	佔總額	人民幣	佔總額	
	千元	百分比	千元	百分比	
	(未經審計)		(未經審計)		
收入					
開放平台及API	825,176	86.5	29,100	15.2	2,735.7
企業級智能體	55,561	5.8	13,739	7.2	304.4
企業級通用大模型	67,037	7.0	147,555	77.3	-54.6
技術服務及其他	6,118	0.7	483	0.3	1,166.7
總計	953,892	100.0	190,877	100.0	399.7

管理層討論及分析

業務回顧

2026年上半年公司在編程(Coding)方面驗證了大模型的可交付性，從早期簡單的編程推薦到完整的代碼文件生成再到一整個工程項目(Agentic Engineering)的實現，本公司完整地走完了這條曲線：模型不再只是一個工具，更是具有獨立把一件事做完的能力。收入結構隨之置換，按調用計費的開放平台與API服務收入佔比由去年全年的26.3%升至86.5%，模型本地化部署收入由73.7%降至13.5%。2026年上半年內主營收入結構完成置換，其實本質是模型能力躍遷之後必然發生的事。網絡安全(Cybersecurity Co-work)是模型同一條能力曲線的延伸。軟件漏洞挖掘不依賴記憶，而是依靠模型的推理深度與長程自主性，這正是編程能力越過工程項目門檻之後自然延伸的方向。報告期內，本公司模型在CyberGym上取得84.5%的成績(超過Fable 5和GPT-5.6 Sol)，在ExploitGym上完成130項任務(僅次於Fable 5和GPT-5.6 Sol)。

支撐這兩條曲線的是基礎設施。本公司自2026年初起全面實施All-in-Infra戰略，在訓練與推理兩側同時投入，並將國產芯片納入主力推理算力，已實現10萬級國產芯片的規模化低成本推理，單位token推理成本較年初下降80%。在全球算力供給持續緊張的條件下，這條路徑決定了本公司的成本曲線掌握在自己手裡。截至報告期末，GLM模型的MaaS平台企業和開發者用戶累計超過580萬。本公司總收入約9.54億人民幣(約1.42億美元)，**同比增長近400%**。其中，開放平台及API業務收入約8.25億人民幣(約1.23億美元)，**同比增長2,735.7%**。

2026年8月，本公司發佈GLM-5.3與GLM-5.3-Flash (Ox-Alpha)，並同步向全量用戶開放GLM Coding Plan訂閱，覆蓋Lite、Pro、Max及團隊版本，這是中國國內最早以訂閱制交付編碼智能體能力的大模型服務之一。截至本公告日期，GLM模型的MaaS平台企業和開發者用戶超過740萬。

一、戰略聚焦

智能上界決定任務邊界，任務價值決定商業空間

2026年初我們提出過一個公式：AGI商業價值=智能上界×Token消耗規模。這次做一點擴展說明：

智能上界決定模型能處理哪些任務、單位token能創造多少用戶價值。更高的智能上界，意味著模型能處理更高價值密度的任務、每單位智能對應更多的經濟產出。Token消耗規模則反映智能被使用的廣度與深度，智能上界先打開任務邊界，任務邊界再決定Token落在什麼價值密度的地方。同樣規模的Token，用於閒聊和用於修復一個生產環境的漏洞，產出相差幾個數量級。

能力階梯。模型能力的演進不是並列的產品線，而是一條只能順序走完的階梯：

Chat → Coding → Agent → Co-work → Autonomous AI

Chat交付一次回答，任務在一輪之內結束，價值密度最低但覆蓋最廣。Coding交付一段可運行的代碼，第一次出現客觀驗收標準—能跑通，或者跑不通。Agent交付一條被執行完的多步任務鏈，模型開始自己調用工具、處理中間失敗。Co-work交付一份可以被專業人士覆核的工作成果，模型進入真實工作流，按完成的任務計價。Autonomous AI交付的是持續運行的能力，模型在無人值守條件下承擔長期目標。每一級向上，都有一個具體的技術門檻把關。從Chat到Coding，門檻是結果可驗證；從Coding到Agent，門檻是長程規劃與錯誤恢復；從Agent到Co-work，門檻是可靠性達到專業人士願意覆核而不是推倒重做的水平；從Co-work到Autonomous AI，門檻是模型能否自己判斷做得對不對。門檻跨不過去，上一層的商業模式就不成立。

市場空間。任務邊界向上，市場空間也逐級放大。按量級粗略估算：Chat對應的搜索市場約1,000億美元，這是互聯網時代給出的答案；Coding對應約5,000億美元的軟件研發支出；Co-work對應約5萬億美元的知識工作開支；更長期看，AI重構操作系統與人機交互方式、帶動各行業升級，對應約30萬億美元的空間。(上述市場空間為本公司基於公開行業數據的粗略估算，僅供說明任務層級與市場量級的對應關係，有關數據未經獨立核證，亦不構成本公司對其業務、財務狀況或經營業績的預測或保證。本公司股東及投資者不應過度依賴該等數據。)相應地，行業的評價標準也在遷移：從單次問答效果和Token消耗量，轉向長程任務成功率、單位智能成本、結果可驗證性，以及客戶的實際經營產出。

報告期內，本公司的重心處在階梯的第二級(Coding)到第四級(Co-work)之間。編程方向已完整走通，網絡安全進入Co-work的驗證階段，法律等方向仍在早期。2026年上半年本公司研發投入約21.31億人民幣(約3.17億美元)，堅持繼續向上攀爬階梯。

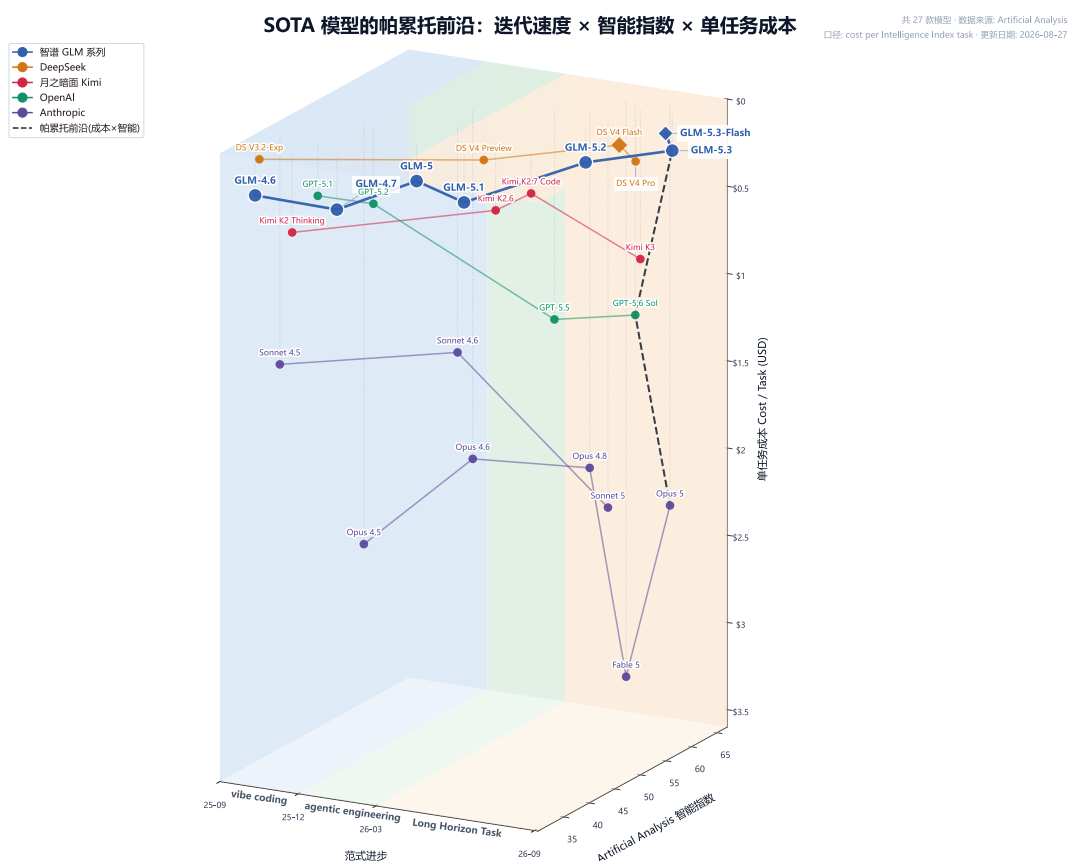
為什麼是這個順序。我們對AGI有一個樸素的**理解**：它不是某一個人的智慧，而是人類智慧水平的總和，最終應當具備產生「相對論」級別原創知識的能力。在這條路上起決定作用的，是模型能力的代際水平。這是本公司所有產品與商業邏輯的起點。

本公司是國內最早投入大模型研發的團隊之一。從GLM預訓練架構、悟道系列到今天的GLM-5系列，我們專注做的**基本**只有一件事：訓練更強的模型。行業熱的時候沒有轉去鋪產品線，行業冷的時候沒有停下迭代。回頭看，收入的每一次快速增長，都發生在模型能力提升、新的任務場景被解鎖之後，而不是之前。

二、技術路徑與模型進展

模型公司的核心能力，是單位資源的智能轉化效率。算力、數據和人都是有價的，如何高效地將它們轉化為智能能力，決定了本公司能走多遠。本公司的極致組織方式就是研究問題從真實任務中來，研究成果直接進入下一代模型，為用戶提供端到端的服務能力。

- 我們如何理解SOTA



過去一年，行業對「SOTA」的理解正在被改寫。時點榜單的領先不再是最重要的事，真正值得關注的是誰能以更快的節奏、更優的成本，持續把更強的模型推向市場。我們選取五家廠商2025年9月至今的27款旗艦模型放進同一個坐標系：橫軸是範式進步，縱軸是智能指數，進深是單任務成本。在這個坐標系裡，單點冠軍只是轉瞬即逝的一個點，持續最優才是一條不斷向上推進的軌跡。各家都在向前走，但節奏與代價不同。**GLM**系列在約11個月內完成4.6、4.7、5、5.1、5.2、5.3六代迭代，智能指數由32提升至60，單任務成本穩定在0.2美元檔，跨越了圖中全部三個範式階段；**GLM-5.3-Flash**以0.045美元的單任務成本，進入智能—成本前沿。（智能指數與單任務成本來源：Artificial Analysis Intelligence Index，取數日期：2026年8月27日，任務口徑：cost per Intelligence Index task。）

報告期內，本公司的技術路徑由三代模型構成。

1. GLM-5：以Coding為入口，打開長程智能

我們最早選擇Coding作為能力突破的入口，是因為軟件工程是最接近「可驗證智能」的真實環境：任務目標清晰，工具鏈完整，代碼能否運行、測試能否通過、問題能否修復，都構成明確的反饋信號。沒有可驗證的反饋，就沒有可優化的目標。

GLM-5的能力重點由代碼生成延伸至系統理解、任務規劃、工具調用、持續調試與工程交付，推動模型從Vibe Coding走向Agentic Engineering。

Coding對本公司的價值不止於幫助開發者寫代碼。在更長的任務鏈路中，模型需要持續理解目標、調用工具、處理異常、根據環境反饋調整策略。通過Coding這一高反饋密度的真實環境積累的長程任務能力，可以遷移到網絡安全、數據分析、複雜自動化及其他專業領域。報告期內，編程方向已成為本公司收入的主要來源，本節前述的能力遷移在網絡安全方向得到驗證。

2. GLM-5.2與GLM-5.3：Scaling deep while scaling up

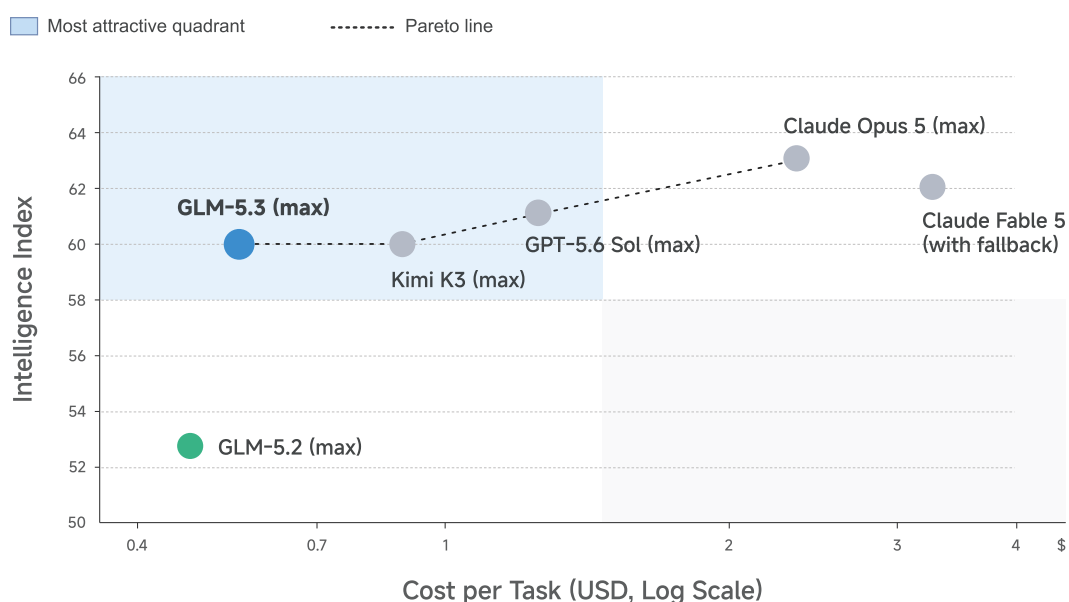
Scaling up是擴大模型規模、上下文、算力與系統承載；Scaling deep是深入後訓練、強化學習、長程軌跡與任務反饋。當前階段，我們把資源優先投向後者。歷史上有過幾次預演。Kaplan定律使參數增速遠超數據增速，Chinchilla以「每參數20 token」的配比修正了這條路徑，70B配1.4T token全面勝過280B的Gopher。進入推理時代，模型每天被調用數十億次，推理成本超過訓練成本，最優點向「更小、訓練更久」移動。MoE進一步把參數拆成兩個量：總參數決定記多少，激活參數與深度決定想多深。近年的研究還表明，最優token/參數比與任務相關，記憶類任務偏好更大參數，推理類任務偏好更多數據。像網絡安全這樣的任務，需要的不是背下更多CVE，而是把二十步因果鏈推到底不散架，這種能力不在總參數裡。

CVE(通用漏洞披露，Common Vulnerations and Exposures)

MoE(混合專家模型，Mixture-of-Experts)

GLM-5.3是這一判斷的一次檢驗。它與GLM-5.2同架構、同總參數、同激活參數，唯一的變量是後訓練規模—用一個月放大長程任務環境與強化學習，在本公司自建的真實現場編碼評測中，端到端完成率較GLM-5.2提升超過50%（評測方法與樣本量詳見GLM-5.3技術博客<https://z.ai/blog/glm-5.3>）。這接近一次受控實驗，說明Scaling不止參數一個因素，而後訓練是當下餘量最大的那一個。

Frontier Models: Intelligence vs. Cost per Task



Source: Artificial Analysis Intelligence Index

我們的思路是：在對的時間，用對的數據，訓練對的規模，才有對的模型。不是簡單的最大，而是在當前數據與算力約束下、把訓練深度做夠之後回報智能上界最高的模型。

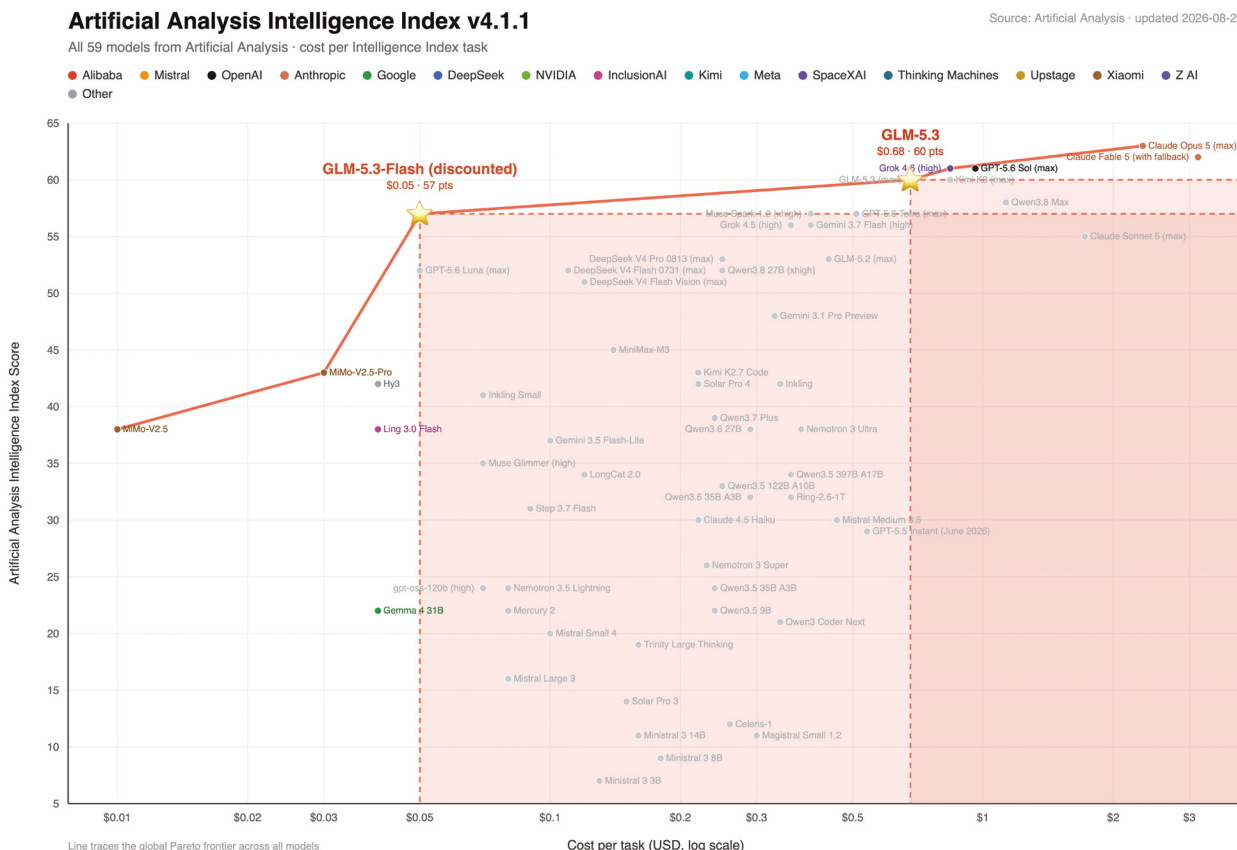
具體展開為四個維度。(i)數據：持續提高知識密度與任務複雜度，通過開發者生態、產業合作與Model-with-Model Self-Play 積累高質量長程Agent軌跡，經篩選、驗證、合成後回流訓練；(ii)中訓練：在預訓練與後訓練之間，本公司單獨配置了中訓練階段的算力，用於提升知識密度、擴展有效上下文長度、建立工具調用與長程任務的基礎模式。三個階段邊際收益不同，算力分開計算，不做一次性投入；(iii)算法：針對長軌跡在獎勵分配、反饋利用與持續學習上的新挑戰，由GRPO探索Critic-based SAO，由SFT向OPD延伸。GLM-5.2起已引入面向超長軌跡的Critic-based單軌跡優化；(iv)工程與環境：Slime提升大規模強化學習訓練效率，Index Share降低長上下文稀疏注意力開銷，MTP優化推理解碼，LayerSplit與ZCube優化KV Cache管理與集群網絡效率。同時持續擴充Coding Agent及更多領域的真實交互環境，讓模型在連續執行、失敗與重規劃中學習。隨著環境數量、類型與複雜度提升，模型學習的已不只是如何回答，而是如何在動態環境中持續完成一項複雜工作。四個維度相互咬合：環境產生軌跡，算法提高軌跡利用率，工程支撐更大規模的訓練與推理，數據決定起點。

3. GLM-5.3-Flash (測試代號Ox-Alpha)：下一代架構與系統能力

如果說GLM-5.3代表當前的能力上界，GLM-5.3-Flash代表的是成本前沿。兩者服務不同的任務區間：前者面向需要最高智能的長程任務，後者面向高頻、大規模、對成本敏感的調用。

GLM-5.3-Flash採用全新架構，專為極低成本設計。總參數320B，激活參數18B，層數45層。它採用稀疏注意力與線性注意力的混合架構，在保持長上下文精度的同時大幅降低長上下文服務成本，並採用流形約束超連接(mHC)提升Scaling能力。結合30T級多模態預訓練語料，GLM-5.3-Flash在基準測試與實際應用中均優於GLM-5.2，價格為後者的十分之一。

正式發佈前，我們以匿名模型Ox-Alpha在OpenCode與OpenRouter上接受真實調用檢驗。上線首日衝至OpenRouter榜首並刷新該平台單日Token用量紀錄；上線一周成為雙平台調用量最大的模型，6天Token調用量超過62萬億，帶動平台整體調用量提升超過20%。



4. 國芯推國模

GLM-5.3-Flash模型的推理是本公司第一次在超大規模真實流量中全面使用國產芯片集群提供服務，芯片之間通過自研高帶寬互聯網絡連接。這項工作的完成方式本身值得記錄：整個推理引擎的構建由GLM-5.3驅動的infra agent大幅加速，它協助工程師開發與優化算子、診斷性能瓶頸、改進部署服務棧。模型優化系統，系統承載模型，這個循環第一次在本公司內部跑通。算子開發週期縮短一半。這也是本公司自身對co-work的一個內部驗證。

技術上的主要約束是國產芯片的內存容量與帶寬，支持1M上下文尤其困難。為此我們在SGLang基礎上構建專用推理引擎，做了以算力換帶寬、以通信換顯存的定製優化，並在集群層面採用生產級Encode-Prefill-Decode分離式架構，將多模態編碼、prompt預填充與逐Token解碼拆成可獨立調度、獨立擴縮容的工作池。與同一硬件上的初始基線相比，端到端服務性能提升3倍；在編碼場景下，單Token的國產芯片推理成本已達到與主流進口卡性價比相當的水平。國產芯片可以在大規模場景下經濟地支撐前沿模型的推理。

三、商業化進展：*Right model will natively go to market*

(一) 收入結構翻轉

本公司的收入結構變化，本質是模型能力提升帶來的交易方式演進：

賣模型→賣調用→賣訂閱→賣端到端任務結果

這四種交易方式不是並列的產品線，而是模型能力提升後依次解鎖的形態。模型能獨立完成的任務越完整，客戶購買的東西就越靠近結果本身，收入的計量單位也隨之改變。

1. 本地化部署階段：完成行業深度積累與驗證

早期模型尚不足以獨立完成一件完整的工作，客戶購買的是工具，而工具需要被集成、被定製、被交付到自有環境中；疊加數據安全、合規與自主可控的要求，本地化部署成為企業採用大模型的自然起點。這一階段的交易方式是賣模型，一次性授權，項目制交付。本公司在這個階段完成了重點行業滲透、複雜場景理解與企業級交付能力沉澱。這些積累今天仍然有效—變化的是交付形態，不是行業理解。

2. Coding階段：從一次性交付轉向持續調用

GLM的編程能力提升推動API與Coding Plan快速增長。本公司在這一階段主動收縮了本地化部署模型軟件授權業務，轉向持續提供可調用、可擴展、可量化的智能服務。2025年底，API調用量較當年6月底增長約7倍，來自開放平台及API的收入佔總收入約26.3%，毛利率提升至18.9%。Coding成為Token消耗與雲端收入增長的第一個規模化入口。這一步的意義不只在增速。收入從一次性確認轉為持續發生，本公司第一次擁有了可預測的經常性收入。

3. Agent與Co-work階段：由模型調用走向任務交付

2026年上半年，GLM在Agent規劃、工具調用、持續執行與長程任務上的能力繼續提升，業務由Coding單點應用向軟件工程、網絡安全、數據分析與複雜自動化等高價值場景延伸。交易方式從賣調用進一步走向賣訂閱、賣端到端的任務結果。

報告期內，開放平台及API業務收入8.25億元，同比增長超過27倍，佔總收入約86.5%，去年同期為15.2%，提升71.3個百分點。

Co-work是這一階段中正在打開的部分：模型連接企業數據、工具與業務系統，嵌入採購、財務、客服、研發、IT運維等環節，從理解需求走到執行任務；企業的真实工作流又反過來成為模型可學習的環境與數據。報告期內本公司已在網絡安全、法律、金融、教育等方向展開嘗試，其中網絡安全進展最快，其餘方向仍處早期，尚未形成規模化收入。

三個階段回頭看是一條線：模型每完成一次能力躍遷，客戶購買的東西就向結果靠近一步，本公司的收入結構隨之改寫一次。

(二) 量價齊升，毛利改善，虧損收窄

上述增長不是以價換量的結果。截至本公告日期，MaaS平台Token調用量較年初增長超過40倍，其中Coding Plan調用量增長超過23倍；同期API平均售價提高約101%，Coding Plan訂閱價格亦有上調。量在漲，價也在漲。

報告期內，開放平台及API業務毛利率提升至24.6%，較去年同期提升約25個百分點，較去年全年提升約6個百分點。本公司期內虧損較去年同期減少12.1%，虧損率(虧損/收入)**較去年同期收窄4.7倍**，經調整期內淨虧損率(經調整期內虧損/收入)**較去年同期收窄3.5倍**。規模擴大的同時單位經濟在改善，這是我們更看重的一組數。

(三) 從Coding走向Co-work

Coding是當前消耗最大、驗證最充分的場景。依託代碼生成、工程理解、工具調用與長程任務能力，本公司由代碼輔助向軟件開發全流程延伸。ZCode通過Harness承擔上下文管理、工具調用、任務調度、緩存與結果校驗，改善模型在真實工程中的表現。隨著長程能力提升，編程任務的複雜度上限還在抬高。

選擇進入哪些Co-work方向，本公司用三個維度判斷：智力門檻，即對推理、規劃、工具調用與長程執行的要求；單位智能消耗，即完成一次任務所需的有效Token、推理時長、算力與人工干預；單位產出價值，即單次交付帶來的收入、節省的人力時間、降低的風險。三者共同決定一項任務今天是否值得交給模型來做。

對大型企業與公共機構，本公司繼續提供本地化部署與組織級智能基礎設施，滿足數據安全、權限管理、系統集成與國產算力適配的要求。這類需求不會消失，但在收入中的角色已從主體變為補充。

(四) 用戶規模與使用深度

截至本公告日期，本公司MaaS平台用戶數超過740萬，較2026年初增長144%；付費日活用戶數較年初增長603%。以收入計的前十大用戶，日均調用量較年初增長98倍。

三個數字的方向一致而幅度遞增：用戶在變多，付費用戶增長更快，已付費的重度用戶用得更深。第三個數字尤其重要，它說明模型正在客戶內部從試用走向生產。

(五) 算力乘數

報告期內，本公司的算力乘數—每投入人民幣1元算力(含訓練與推理)所對應的開放平台及API收入—較去年同期提升約14倍。

這條曲線的改善來自兩頭：單位Token承載的智能在提高，單位任務消耗的Token與成本在下降。前者由模型能力決定，後者由基礎設施決定；兩者同時改進，才有「智能—成本」的帕累托推進。每投入人民幣1元算力換回的收入在增加，意味著模型的自我造血能力在增強。我們希望把這條曲線繼續往上推。

四、生態與安全：讓能力獲得真實世界的信任

模型能力要變成可依賴的服務，中間隔著一件事：真實世界是否信任它。本公司在這個方向上同時做兩件事：開源開放，讓安全能力與模型能力同步增長。一方面，本公司通過開放模型權重、MaaS平台、Coding Plan與全球開發者網絡，讓模型進入盡可能多的真實環境。另一方面，本公司通過權限控制、過程監督、風險評估、漏洞披露與外部驗證，建立能力的可核查性。兩件事互為條件：沒有開源開放，安全無從檢驗；沒有安全，開源開放無法持續。

安全不只是合規成本，它也可以是一種可驗證的公共品，幫助市場判斷一個模型能否被安全、穩定、長期地部署到真實環境中。

從能力到公共品

自GLM-5.2發佈以來，本公司聯合多家安全團隊開展了持續的紅隊測試與安全評估。截至報告期末，累計發現漏洞2,436個(經初篩與去重)，其中，中高危1,097個，覆蓋系統內核、操作系統、瀏覽器引擎、開源基礎組件、互聯網應用與網絡協議等269個項目。其中相當一部分是長期存續、未被發現的歷史漏洞，最長的已存在20餘年。

這些發現已經形成一項持續的漏洞發現與披露工作，相關漏洞陸續提交至CNNVD/CNVD等國家漏洞庫。為提高透明度，本公司建立了安全披露賬本(<https://cvd.z.ai/>)，記錄漏洞從發現、確認到修復的全過程。賬本持續更新，僅公開已完成披露並由廠商修復的漏洞；對仍處於協調披露階段的漏洞，公開哈希值以供後續核驗，不公開可復現的利用代碼。

這套機制的意義在於，它把模型的能力放在一個第三方可以核驗的位置上。基準分數由我們自己報告，漏洞披露則由國家漏洞庫和廠商確認(後者是外部信任的來源)。

開源與開放

隨GLM-5.3發佈，本公司同步啟動「開源的盾」計劃，包含三項持續投入：第一，對重點開源項目開展持續安全審計，幫助維護者免費發現並修復風險，審計結果通過OpenVuln公開(<https://huggingface.co/spaces/zai-org/OpenVuln>)。第二，向開源社區免費提供模型額度與防禦入口，開源項目維護者可申請用於安全審計和防禦任務。第三，在ZCode中提供代碼審計功能，使安全檢查進入日常研發流程。

這三件事的共同點是把本公司最前沿的能力，交給最需要它但最付不起錢的一方。開源基礎組件支撐著幾乎全部現代軟件，維護者往往是無償的少數人；攻擊方使用前沿模型的成本正在快速下降，防守方卻未必負擔得起。當最強的矛被鎖在少數人手裡，最好的盾必須屬於所有人。在開源之前，本公司聯合國內頂級安全團隊開展持續紅隊測試與自動化越獄評估，用新產生的樣本迭代加固審查系統，並邀請專業團隊對模型的高風險任務能力進行獨立評估。能力釋放與風險防護之間的平衡，則由外部機構而非本公司自己判斷。

從商業角度看，開源與安全並不是成本項。開源帶來的分發規模，是模型獲得真實任務反饋的前提；安全披露帶來的外部確認，是企業客戶決定是否把模型接入生產系統的依據。

能力邊界與受控開放

模型在網絡安全方向上的能力具有雙重用途。本公司對最敏感的能力採取受控開放：面向經過驗證的安全研究者與機構提供，其餘用戶按標準安全策略使用。在模型開源前，本公司完成安全評估、模型加固與使用邊界設計。

未來展望

2026年上半年我們驗證了模型可以獨立完成一件完整的工作。下一步，是讓它承擔一段完整的業務。

一、任務：從長程任務到自治系統

更長的時間跨度。讓模型在更長的週期內保持目標一致，持續進行任務拆解、工具調用、環境交互、錯誤修復與結果驗證。今天的前沿是數小時量級的工程任務，下一步是數天量級的完整交付。

多智能體協同。從單一Agent走向多Agent分工協作，由模型承擔完整業務流程，從輔助工作走向執行工作。這一步的門檻不完全在智能本身，更在可靠性與可審計—模型要能在無人逐步檢查的條件下工作，同時留下可覆核的痕跡。

任務邊界每向外推一層，對模型能力的要求就換一個量級。因此接下來的問題是：能力還能沿哪些維度繼續提升。

二、能力：Scaling從未停止

Scaling依然是通向更高智能的主路徑，這一點沒有變，變的是Scaling什麼。

過去幾年，Scaling幾乎等同於參數量。但參數越過「裝下世界」的閾值之後，繼續加寬的邊際收益迅速下降，而其他維度的餘量還遠未用盡。今天決定一個模型上限的，是四個因素的組合：基座規模、有效深度、訓練深度、任務環境。四者邊際收益不同，且隨階段輪換；本公司的資源分配，就是在不同階段判斷哪個因素當前回報最高。

基座規模：上限由預訓練決定。後訓練是當期回報最高的旋鈕，但智能的天花板由基座決定。一個模型能被訓練到多深，取決於它在預訓練階段見過多少世界、留下了多少可被激活的結構。把後訓練做到極致之後，下一段增長必須回到基座。本公司從未停止基座迭代—GLM-5.3-Flash已使用新一代架構與30T級多模態預訓練語料，驗證了新架構在中等參數量級上的效率；在此基礎上，本公司正在推進下一代基座模型的訓練，方向包括更大的有效規模、更長的原生上下文與原生多模態的統一建模。

有效深度：讓模型多想幾步。參數決定記多少，深度決定想多深。循環架構(Loop Transformer)通過層的循環復用，在不顯著增加參數量的前提下提高有效計算深度，把算力從「記」轉向「想」(本公司已在初步實驗中驗證了這一想法)。這條路徑與我們對任務的判斷一致：網絡安全這類工作需要的是把二十步因果鏈推到底不散架，這種能力來自深度，不來自總參數。

訓練深度：當前餘量最大的一個。GLM-5.3是這一判斷的直接證據—與GLM-5.2同架構、同總參數、同激活參數，唯一變量是後訓練規模，端到端完成率提升超過50%。這條曲線尚未見頂，本公司將繼續擴大長程任務的強化學習規模、延長有效訓練軌跡、提高軌跡利用效率。

任務環境：新出現的變量。當算法與算力都不再是唯一瓶頸，環境的數量、類型與複雜度開始直接決定模型能學到什麼。這是過去兩年新出現的維度，也是本公司下一階段投入最集中的方向。

三、系統：*Fully Self Training*

四個旋鈕中，環境的擴建速度正在成為新的瓶頸。而更深一層的變化是：訓練系統的三個組成部分：數據、環境、基礎設施，都出現了由AI接管的早期證據。

數據開始自產。高質量人類數據正在見頂，公域數據接近用盡，而下一代基座需要的是數以萬億計、質量只升不降的token。我們的解法是一條model-vs-model的自我對弈管線：一個模型生成解法與內容，另一個模型評審、測試、執行驗證，通過對抗檢驗的數據才進入下一環節。之後是四層過濾—規則校驗確認可編譯、推導自洽、格式合規；執行驗證在沙盒中實際運行，執行結果本身就是最高質量的訓練信號；模型評審再加一道判斷；最後人工抽檢校準。沉澱下來的是可驗證、可執行、可復盤的數據，同時反哺預訓練、中訓練與後訓練三個階段。

因此，數據的質量上限不再取決於人類標注的速度，而取決於模型自身的驗證能力。模型越強，生成的問題越難、驗證越嚴，數據質量越高；更好的數據再訓練出更強的下一代。數據從一次性的消耗品，變成可跨代復用的資本品。

環境開始自造。一個有用的訓練環境必須可執行、可驗證、貼近真實專業工作，而且需要成千上萬個，手工搭建供不上。在GLM-5.3的技術博客中，我們分享了一些我們的做法：研究智能體從真實工作中採集任務模式，把資深工程師數天工作量的任務轉化為可運行的長程環境，內含多步依賴與隱藏狀態；裁判智能體逐一試做，確認任務確實可解；驗證器在無法訪問參考答案的條件下被合成，再用求解器軌跡發現並封堵獎勵捷徑；通過oracle、no-op、未解狀態三項檢查的驗證器，最終產出可直接用於強化學習的二值獎勵。slime架構把數學、代碼、沙盒、驗證器降格為即插即用的數據組件，環境成為流水線上的標準產品。

需要說明的是，這條管線目前仍有相當數量的人工環節。讓環境的生成與驗證更加自主，是我們的下一步。

基礎設施開始自我優化。推理系統的優化：內核、調度、量化、算子等，本質上是代碼工程任務，而這恰是模型最強的能力域。GLM-5.3-Flash的推理引擎就是這樣完成的：由GLM-5.3驅動的基礎設施智能體協助工程師開發優化內核、診斷性能瓶頸、改進服務棧，最終在國產芯片上把端到端服務性能提升三倍。同一邏輯也在訓練側運轉：slime讓訓練與大規模推理rollout共用一套數據流，強化學習中積累的配置經驗與調度策略可直接遷移到生產服務階段。

三條線指向同一個終局：模型參與自身的改進，形成遞歸式的自我改進(Recursive Self-Improvement)。這不是一個遙遠的概念，2026年上半年本公司已經出現了它的早期形態：GLM-5.3參與了服務GLM自身的推理引擎優化，GLM生成的環境正在訓練下一代GLM。差別只在於，今天這個迴路裡還站著大量工程師。迴路每收緊一環，智能進步的速度就更多地由環境生長的速度決定。

這就是我們的技術願景：Fully Self Training，下一代GLM，將在上一代GLM搭建的環境裡訓練。

四、邊界與治理

本公司的技術願望是模型可以自動參與訓練系統的建設，但也需要保證對其能力的評估與風險的判斷最終由人來裁量，並引入外部機構獨立評估。2026年上半年我們在網絡安全方向上已經這樣做了，即最敏感的能力受控開放，開源前由第三方專業團隊獨立評估，漏洞通過國家漏洞庫完成責任披露。基準分數由我們自己報告，風險判斷交給外部確認，這個分工不會因為能力提升而改變，只會隨能力上移而加強。

本公司將繼續在能力上限、推理效率、任務價值密度與安全治理之間尋找新的平衡，讓更高水平的智能穩定、可控地進入真實世界，讓全人類享受普惠智能。

財務回顧

以下討論乃基於本中期業績公告其他部分所載財務資料及相應附註，並應與之一併閱讀。

經營業績主要項目分析

收入

報告期內，我們實現收入大幅增長。截至2025年6月30日止六個月及截至2026年6月30日止六個月，我們的收入分別為人民幣190.9百萬元及人民幣953.9百萬元，增幅達399.7%，收入規模的擴張主要得益於雲端部署業務的爆發式增長，標誌著本集團收入增長的主要驅動由本地化部署業務轉向雲端部署業務。

按部署方式來看：

雲端部署

雲端部署收入由截至2025年6月30日止六個月的人民幣29.1百萬元增加至截至2026年6月30日止六個月的人民幣825.2百萬元，增幅達2,735.7%。該增長主要反映本公司模型能力逐步面向高價值任務場景提升，模型長程任務能力的持續進步，讓用戶能夠在複雜場景下持續調用，進而帶來用戶日均Token調用顯著提升；MaaS平台用戶規模持續擴大，量價齊升態勢延續。雲端部署收入佔總收入的比例由上年同期的15.2%提升至86.5%，成為本集團主要的收入來源。

本地化部署

本地化部署收入由截至2025年6月30日止六個月的人民幣161.8百萬元下降至截至2026年6月30日止六個月的人民幣128.7百萬元，減幅達20.5%。該減少主要由於本集團在雲端部署業務取得爆發式增長的情況下，更強調本地化業務的質量和行業選擇，夯實未來模型能力從泛化通用場景進入垂直行業高價值特化任務場景的基礎。

截至6月30日止六個月

	2026年		2025年	
	人民幣千元 (未經審計)	佔總額 百分比	人民幣千元 (經審計)	佔總額 百分比
雲端部署	825,176	86.5%	29,100	15.2%
本地化部署	128,716	13.5%	161,777	84.8%
總計	<u>953,892</u>	<u>100.0%</u>	<u>190,877</u>	<u>100.0%</u>

按業務形態與核心產品線來看：

開放平台及API

開放平台及API(本集團基於通用大模型提供的雲端服務)收入由截至2025年6月30日止六個月的人民幣29.1百萬元增加至截至2026年6月30日止六個月的人民幣825.2百萬元，增幅達2,735.7%，該增長主要反映本公司模型能力逐步面向高價值任務場景提升，GLM系列模型多輪迭代帶動長程任務能力的持續進步與質變，讓用戶能夠在複雜場景下持續調用，進而帶來用戶日均Token調用顯著提升；MaaS平台用戶規模持續擴大，量價齊升態勢延續。

企業級智能體

企業級智能體(本集團面向企業複雜場景構建的自主智能系統產品)收入由截至2025年6月30日止六個月的人民幣13.7百萬元增加至截至2026年6月30日止六個月的人民幣55.6百萬元，增幅達304.4%，得益於企業客戶對智能體的採納加速：一方面，GLM系列模型在長程任務與複雜場景下的執行能力持續提升，使得面向企業工作流的智能體產品得以覆蓋更多高價值場景；另一方面，報告期內，企業級客戶對智能體場景的部署需求持續增加，使得該分部收入實現增長。

企業級通用大模型

企業級通用大模型(本集團自主研發的預訓練模型矩陣，主要採用私有化本地部署模式交付)收入由截至2025年6月30日止六個月的人民幣147.6百萬元減少至截至2026年6月30日止六個月的人民幣67.0百萬元，減幅達54.6%。該減少主要由於本集團更強調企業級通用大模型業務的質量和行業選擇，夯實未來模型能力從泛化通用場景進入垂直行業高價值特化任務場景的基礎；同時部分企業級通用大模型客戶逐步向開放平台及API服務轉化，形成了正飛輪。

技術服務及其他

技術服務及其他收入由截至2025年6月30日止六個月的人民幣0.5百萬元增加至截至2026年6月30日止六個月的人民幣6.1百萬元，較上年同期增加1,120.0%，主要由於本集團將自研大模型能力向更多場景進行了延伸探索。該分部收入基數較小，其變動對本集團整體收入結構的影響有限。

	截至6月30日止六個月			
	2026年		2025年	
	人民幣千元 (未經審計)	佔總額 百分比	人民幣千元 (未經審計)	佔總額 百分比
開放平台及API	825,176	86.5%	29,100	15.2%
企業級智能體	55,561	5.8%	13,739	7.2%
企業級通用大模型	67,037	7.0%	147,555	77.3%
技術服務及其他	6,118	0.7%	483	0.3%
總計	953,892	100.0%	190,877	100.0%

毛利和毛利率

本集團毛利由截至2025年6月30日止六個月的人民幣95.4百萬元增加至截至2026年6月30日止六個月的人民幣251.6百萬元，增幅達163.7%。本集團毛利率由截至2025年6月30日止六個月的50.0%減少至截至2026年6月30日止六個月的26.4%，主要由於收入結構轉變所致，雲端部署佔比急速擴大，該類業務尚在爆發初期，毛利率尚在爬坡，結構性攤薄了整體毛利率。

但雲端業務自身毛利率已由去年同期的負數轉為正數，並快速提升，報告期內，本集團毛利實現了逾1.6倍的增長，反映收入結構變化已形成正向毛利貢獻。

按部署方式來看：

雲端部署

雲端部署業務的毛利由截至2025年6月30日止六個月的人民幣-0.1百萬元增加至截至2026年6月30日止六個月的人民幣202.9百萬元，較上年同期由負轉正。毛利率由截至2025年6月30日止六個月的-0.4%增加至截至2026年6月30日止六個月的24.6%。該改善主要得益於：其一，收入規模爆發式增長帶來的規模效應，指數級增長的Token調用量顯著攤薄了固定成本；其二，定價策略隨模型智能上界的提升逐步上調，但調用量依然持續增長，維持量價齊升態勢；其三，全棧自研的工程優化持續降低模型推理成本。

本地化部署

本地化部署業務的毛利由截至2025年6月30日止六個月的人民幣95.5百萬元下降至截至2026年6月30日止六個月的人民幣48.7百萬元，較上年同期減少49.0%。毛利率由截至2025年6月30日止六個月的59.1%下降至截至2026年6月30日止六個月的37.9%。毛利及毛利率均有所下降，主要由於本集團推進業務模式調整，本地化部署的產品結構發生變化，毛利率較高的企業級通用大模型私有化交付規模收縮，而尚處於規模化拓展階段的企業級智能體產品佔比提升；同時部分客戶由本地化部署向雲端服務遷移。

截至6月30日止六個月

	2026年		2025年	
	毛利	毛利率	毛利	毛利率
	(人民幣千元)	(%)	(人民幣千元)	(%)
	(未經審計)		(經審計)	
雲端部署	202,871	24.6	(116)	-0.4
本地化部署	48,740	37.9	95,540	59.1
總計	<u>251,611</u>	<u>26.4</u>	<u>95,424</u>	<u>50.0</u>

按業務形態與核心產品線來看：

開放平台及API

開放平台及API業務的毛利由截至2025年6月30日止六個月的人民幣-0.1百萬元增加至截至2026年6月30日止六個月的人民幣202.9百萬元，較上年同期由負轉正。毛利率由截至2025年6月30日止六個月的-0.4%改善至截至2026年6月30日止六個月的24.6%。該改善主要得益於規模效應的顯現、定價的上調以及全棧工程優化對推理成本的持續壓低。

企業級智能體

企業級智能體業務的毛利由截至2025年6月30日止六個月的人民幣8.9百萬元增加至截至2026年6月30日止六個月的人民幣19.8百萬元，較上年同期增加122.5%。毛利率由截至2025年6月30日止六個月的64.6%減少至截至2026年6月30日止六個月的35.6%。毛利增加但毛利率有所下降，主要由於報告期內企業級客戶對智能體場景的部署需求日益殷切，需求持續增加，為保障交付質量與場景覆蓋深度，前期為滿足客戶多元化需求投入較大；隨著規模化效應的逐步顯現，毛利率有望企穩回升。

企業級通用大模型

企業級通用大模型業務的毛利由截至2025年6月30日止六個月的人民幣86.3百萬元減少至截至2026年6月30日止六個月的人民幣28.4百萬元，較上年同期減少67.1%。毛利率由截至2025年6月30日止六個月的58.5%減少至截至2026年6月30日止六個月的42.3%。毛利及毛利率均有所減少，主要由於本集團更強調該類業務的質量和行業選擇，夯實未來模型能力從泛化通用場景進入垂直行業高價值特化任務場景的基礎，而該類業務交付及維護成本具備一定剛性，規模效應減弱；同時客戶需求升級，部分項目轉向企業級智能體方案，亦對毛利率產生階段性影響。

技術服務及其他

技術服務及其他業務的毛利由截至2025年6月30日止六個月的人民幣0.3百萬元增加至截至2026年6月30日止六個月的人民幣0.6百萬元，較上年同期增加100.0%。毛利率由截至2025年6月30日止六個月的65.8%減少至截至2026年6月30日止六個月的9.4%。毛利增加惟毛利率有所下降，主要歸因於報告期內部分探索性場景項目投入增多，且因該分部收入基數較小，毛利率影響較為顯著。

	截至6月30日止六個月			
	2026年		2025年	
	毛利	毛利率	毛利	毛利率
	(人民幣千元)	(%)	(人民幣千元)	(%)
	(未經審計)		(未經審計)	
開放平台及API	202,871	24.6	-116	-0.4
企業級智能體	19,796	35.6	8,880	64.6
企業級通用大模型	28,368	42.3	86,342	58.5
技術服務及其他	576	9.4	318	65.8
總計	251,611	26.4	95,424	50.0

銷售成本

本集團的銷售成本由截至2025年6月30日止六個月的人民幣95.5百萬元增加至截至2026年6月30日止六個月的人民幣702.3百萬元，較上年同期增加635.4%。成本增加主要歸因於配合業務擴張及收入增長所導致的計算服務費用增加，與收入增長趨勢基本一致。

資本支出

本集團的資本支出由截至2025年6月30日止六個月的人民幣53.7百萬元增加至截至2026年6月30日止六個月的人民幣390.6百萬元，較上年同期增長627.4%。該增長主要由於本公司完成收購北京紅鑽科技發展有限公司（「北京紅鑽」）全部股權，新增大額物業、廠房及設備。

其他收入

本集團的其他收入由截至2025年6月30日止六個月的人民幣4.6百萬元增加至截至2026年6月30日止六個月的人民幣41.4百萬元，較上年同期增加800.0%。此乃主要由於本集團於2026年1月上市後銀行存款規模大幅增加，令利息收入相應上升。

研發開支

本集團的研發開支由截至2025年6月30日止六個月的人民幣1,594.7百萬元增加至截至2026年6月30日止六個月的人民幣2,131.2百萬元，較上年同期增加33.6%。該增長主要由於本集團持續加大模型訓練的研發投入，以支撐GLM系列模型智能上界的持續突破。

銷售及營銷開支

本集團的銷售及營銷開支由截至2025年6月30日止六個月的人民幣208.6百萬元減少至截至2026年6月30日止六個月的人民幣177.7百萬元，較上年同期減少14.8%。此乃本集團優化市場推廣資源投放、提升營銷投入的轉化效率所致。

一般及行政開支

本集團的一般及行政開支由截至2025年6月30日止六個月的人民幣185.2百萬元減少至截至2026年6月30日止六個月的人民幣103.3百萬元，較上年同期減少44.2%。該減少主要是由於報告期內專業服務費以及以權益結算以股份為基礎的薪酬開支較上年同期減少。

金融資產減值虧損

本集團的金融資產減值虧損由截至2025年6月30日止六個月的人民幣10.9百萬元增加至截至2026年6月30日止六個月的人民幣27.4百萬元，較上年同期增加151.4%。此乃主要由於業務規模擴大導致貿易應收款項及應收票據原值增加。

財務成本

本集團的財務成本由截至2025年6月30日止六個月的人民幣53.3百萬元增加至截至2026年6月30日止六個月的人民幣95.9百萬元，較上年同期增加79.9%。此乃主要由於報告期內錄得淨匯兌虧損及銀行貸款利息開支增加。

分佔聯營公司利潤減虧損

本集團分佔聯營公司利潤減虧損由截至2025年6月30日止六個月的人民幣14.1百萬元增加至截至2026年6月30日止六個月的人民幣144.2百萬元，較上年同期增加922.7%。該增加主要歸因於聯營公司報告期內經營及財務業績改善。

按公允價值計入損益的金融工具的公允價值變動

按公允價值計入損益計量的金融工具的公允價值變動由截至2025年6月30日止六個月的人民幣9.8百萬元增加至截至2026年6月30日止六個月的人民幣50.4百萬元，較上年同期增加414.3%。此乃主要由於本集團持有的金融工具於期內估值上升，確認公允價值收益所致。

向投資者發行的金融工具的賬面值變動

向投資者發行的金融工具的賬面值由截至2025年6月30日止六個月虧損人民幣429.3百萬元減少至截至2026年6月30日止六個月虧損人民幣22.1百萬元，較上年同期減少94.9%。此乃主要由於向投資者發行的金融工具及相關贖回權終止，該等金融工具已於2026年1月上市時終止確認，報告期內僅就上市前的短暫期間確認賬面值變動所致。

虧損淨額

截至2026年6月30日止六個月，本集團錄得淨虧損人民幣2,072.0百萬元，較截至2025年6月30日止六個月的淨虧損人民幣2,357.9百萬元有所減少。主要由於報告期內毛利增加以及上市後向投資者發行的金融工具已轉換為普通股，相關賬面值變動虧損大幅減少；研發開支增加則抵銷了部分減幅。

非國際財務報告準則計量

	截至6月30日止六個月	
	2026年	2025年
	人民幣千元	人民幣千元
	(未經審計)	(未經審計)
期內虧損	(2,071,992)	(2,357,852)
加：		
—以權益結算以股份為基礎的薪酬開支	85,764	158,852
—向投資者發行的金融工具賬面值變動	22,090	429,295
—上市開支	—	17,731
經調整期內淨虧損(非國際財務報告準則計量)	(1,964,138)	(1,751,974)

財務狀況主要項目分析

財務狀況

股東權益由2025年12月31日的人民幣-8,111.0百萬元增加至2026年6月30日的人民幣4,429.6百萬元，主要由於本集團於報告期內成功上市，向投資者發行的金融工具及相關贖回權已終止，且有關金融負債已轉入本公司權益。

流動性及財務資源

於2026年6月30日，本集團持有現金及現金等價物人民幣3,993.7百萬元，較2025年12月31日的人民幣2,259.1百萬元增加人民幣1,734.6百萬元。此波動主要歸因於報告期內收到上市募集資金並增加銀行貸款，同時營運開支消耗了部分資金。

債務/借款

於2026年6月30日，本集團銀行貸款總額為人民幣2,224.8百萬元，較2025年12月31日的人民幣689.5百萬元增加人民幣1,535.3百萬元，主要為滿足經營需要而借入資金。相關貸款均為帶息借貸，主要以人民幣為計價貨幣，上述借貸均為無擔保借貸。

租賃負債

於2026年6月30日，本集團確認的租賃負債總額(包括流動及非流動部分)為人民幣379.4百萬元，較2025年12月31日的人民幣545.1百萬元減少人民幣165.7百萬元。該波動主要歸因於報告期內租賃款項的支付以及租賃安排的變動。

資本負債比率

本集團於2026年6月30日的資本負債比率(資本負債比率以銀行貸款及租賃負債除以截至期末權益總額再乘以100%計算)為58.8%，而於2025年12月31日為-15.2%。該變動主要由於上市時向投資者發行的金融工具及相關贖回權終止，且有關金融負債已轉入本公司權益，令權益總額由負數轉為正數所致。

資產抵押

截至2026年6月30日止六個月，本集團資產均無任何抵押。

資本承擔

截至2026年6月30日止六個月，本集團無任何未履行的資本承擔。

或有負債

截至2026年6月30日止六個月，本集團並無任何重大或有負債。

重大投資及附屬公司、聯營公司及合營企業的重大收購及處置

根據日期為2026年3月31日的董事會決議及日期為2026年4月14日的收購框架協議，本公司以不超過人民幣360.52百萬元的總代價收購北京紅鑽的100%股權。於收購事項完成後，北京紅鑽已作為本公司一間附屬公司列賬。詳情請參閱本公司日期為2026年4月14日之公告。

除上文所披露者外，截至2026年6月30日，本集團並未進行任何重大投資，亦無任何其他有關附屬公司、聯營公司及合營企業的重大收購或出售事項；截至本公告日期，本集團並無購入重大資本資產的具體未來計劃。

僱員及薪酬政策

於2026年6月30日，本集團有981名正式僱員(2025年6月30日：883名)，本集團聘用的僱員人數視乎需要而進行變動。於報告期內，本集團產生總薪酬成本(含股份支付)為人民幣482.1百萬元(截至2025年6月30日止六個月：人民幣499.6百萬元)。

人才是本公司最核心的戰略資源。我們圍繞人工智能行業的特點，打造了以能力為導向、以激勵為驅動的人才管理體系，支撐本公司在AI領域的持續技術創新與業務擴張。

在薪酬策略上，我們採用「固定薪酬+績效獎金+中長期激勵」的三層結構，確保回報與本公司業績及個人貢獻深度綁定。通過員工激勵計劃及持股平台，將核心管理人員、核心技術人員及業務人才的利益與公司長期發展緊密對齊。我們定期開展行業薪酬對標，保持薪酬水平的市場競爭力，以吸引和留住頂尖人才。

在員工關懷方面，我們在法定社會保險及住房公積金的基礎上，提供補充商業保險、年度健康體檢、餐飲補貼及節日福利等多層次保障。同時支持各類文體社團活動，關注員工身心健康，營造積極向上的工作氛圍。本公司建立了系統化的內部培訓與知識共享機制，鼓勵跨部門協作與技術交流，為員工的職業發展提供持續支持。

作為大模型領域的領軍企業，我們將AI原生工作模式確立為組織運行的基本方式。以自研模型及智能體系統為底座，我們將自身打造為模型能力在真實生產環境中規模化驗證的前沿場域。在日常研發中，代碼生成、缺陷修復、測試編寫等標準化環節多數由AI自主完成，員工則專注於問題本質的提煉、技術方案的設計與創新方向的開拓。這套「AI幹活、人做決策」的協作範式，正在重塑本公司的產出效率與創新節奏。

匯率波動風險

本集團的外匯風險來自未來商業交易以及以相關集團實體的功能貨幣以外的貨幣計值的已確認資產及負債。本公司的功能貨幣為人民幣。大部分非人民幣資產及負債為以美元及港元計值的現金及現金等價物。

本集團主要面臨人民幣兌美元與人民幣兌港元匯率變動的風險。我們目前並無進行旨在或擬用於管理外匯匯率風險的對沖活動，然而，我們透過密切監察外幣匯率變動，以管理其外匯風險並將採取必要措施減低此等風險。

主要風險及不明朗因素

我們的運營涉及若干風險及不確定因素，其中包括(i)與我們研發相關的風險；(ii)與我們商業化相關的風險；(iii)與我們運營相關的風險；(iv)與我們知識產權相關的風險；(v)與我們的財務狀況及額外資金需求相關的風險；及(vi)與我們開展業務所在司法管轄區有關的風險，包括但不限於：

- 人工智能行業屬於前沿科技領域，具有持續變化的特點。本集團的未來業務、經營業績、財務狀況及競爭地位取決於本集團對模型和技術不斷升級反覆運算的能力。
- 我們已經並預計將繼續在研發方面投入大量資金，不斷更新反覆運算大模型的研發工作。
- 本集團面臨客戶變化及新客戶拓展不及預期的風險。
- AGI的開發仍處於早期階段，未來實現AGI存在很大的不確定性。
- 我們需長期維持充足、穩定的技術人才隊伍以及時應對市場變化。若核心技術人員發生不利變化，將會影響本集團重要研發項目的開展、推進及商業化，進而可能對本集團經營產生不利影響。
- 我們面臨規模擴張帶來的管理風險。
- 本集團資產、營收規模呈快速增長，人員數量持續增加，對於本集團的管理水準和風險控制提出更高的要求。本集團可能在一定程度上面臨規模擴張所帶來的管理風險。
- 我們依賴第三方為我們提供計算資源，其服務的任何中斷或價格波動均可能對我們的業務、經營業績及財務狀況產生不利影響。

- 中國的人工智能行業正處於不斷發展且日益嚴格的監管環境中。未來可能出台的法律法規或將提出更多要求與其他義務，這可能對我們的業務、經營業績、財務狀況及前景產生不利影響。
- 我們可能無法為我們的業務獲得或維持足夠的知識產權保護，或者該等知識產權保護的範圍可能不夠廣泛。
- 我們面臨未來市場競爭加劇的風險。未來若本集團技術發展和產品反覆運算無法保持在行業領先水準，本集團的競爭地位、市場份額和經營業績可能受到不利影響。
- 受多種因素影響，我們目標市場的規模及對我們解決方案的需求可能無法按我們預期的速度增長，這將對我們的業務、經營業績、財務狀況及前景產生不利影響。
- 若我們的MaaS平台未能按要求運行，可能損害我們的業務、經營業績、財務狀況及前景。
- 我們在業務商業化方面的往績記錄有限，且大模型市場同時受到技術反覆運算和監管政策變動的雙重影響，發展機遇與未來市場規模難以準確預測。

以上所列並非全部，具體請參照我們的招股章程「風險因素」章節。

綜合損益及其他全面收益表

截至2026年6月30日止六個月－未經審計
(以人民幣列示)

	附註	截至6月30日止六個月	
		2026年 人民幣千元	2025年 人民幣千元
收入	3	953,892	190,877
營業成本		<u>(702,281)</u>	<u>(95,453)</u>
毛利	3(b)	251,611	95,424
其他收入		41,436	4,614
銷售及營銷開支		(177,666)	(208,570)
一般及行政開支		(103,346)	(185,165)
研發開支		(2,131,171)	(1,594,661)
金融資產減值虧損		<u>(27,416)</u>	<u>(10,867)</u>
經營虧損		(2,146,552)	(1,899,225)
融資成本		(95,925)	(53,270)
分佔聯營公司利潤減虧損		144,152	14,147
按公允價值計入損益(「按公允價值計入損益」)的金融工具的公允價值變動		50,385	9,791
向投資者發行的金融工具的賬面值變動		<u>(22,090)</u>	<u>(429,295)</u>
除稅前虧損		(2,070,030)	(2,357,852)
所得稅	4	<u>(1,962)</u>	<u>—</u>
期內虧損		<u><u>(2,071,992)</u></u>	<u><u>(2,357,852)</u></u>

綜合損益及其他全面收益表

截至2026年6月30日止六個月－未經審計
(以人民幣列示)

		截至6月30日止六個月	
		2026年	2025年
	附註	人民幣千元	人民幣千元
期內虧損		<u>(2,071,992)</u>	<u>(2,357,852)</u>
期內其他全面收益(除稅後)：			
可能重新分類至損益的項目：			
－將財務報表換算為列報貨幣的 匯兌差額		<u>(3,049)</u>	<u>414</u>
期內全面收益總額		<u>(2,075,041)</u>	<u>(2,357,438)</u>
以下人士應佔虧損：			
本公司權益股東		(2,071,326)	(2,351,173)
非控股權益		<u>(666)</u>	<u>(6,679)</u>
		<u>(2,071,992)</u>	<u>(2,357,852)</u>
以下人士應佔全面收益總額：			
本公司權益股東		(2,074,375)	(2,350,759)
非控股權益		<u>(666)</u>	<u>(6,679)</u>
		<u>(2,075,041)</u>	<u>(2,357,438)</u>
每股虧損			
基本及攤薄(人民幣元)	5	<u>(4.67)</u>	<u>(6.23)</u>

綜合財務狀況表

於2026年6月30日－未經審計
(以人民幣列示)

		2026年 6月30日 人民幣千元	2025年 12月31日 人民幣千元
	附註		
非流動資產			
物業、廠房及設備	6	855,060	655,825
無形資產		67,801	50,963
商譽		283,815	39,379
於聯營公司的權益		532,133	338,081
其他非流動資產		912,325	198,202
		<u>2,651,134</u>	<u>1,282,450</u>
流動資產			
按公允價值計入損益的短期投資		506,139	377,287
存貨及合約成本		87,100	125,817
貿易應收款項及應收票據	7	782,944	303,208
預付款項及其他應收款項	8	1,621,311	395,734
合約資產		2,480	1,625
定期存款		—	108,593
銀行存款及手頭現金		3,993,722	2,259,147
		<u>6,993,696</u>	<u>3,571,411</u>
流動負債			
貿易及其他應付款項	9	2,135,759	1,328,631
合約負債		302,325	148,644
銀行貸款		789,831	604,836
租賃負債		161,711	251,316
向投資者發行的金融工具		—	10,072,827
		<u>3,389,626</u>	<u>12,406,254</u>

綜合財務狀況表

於2026年6月30日－未經審計
(以人民幣列示)

	2026年 6月30日 人民幣千元	2025年 12月31日 人民幣千元
流動資產/(負債)淨額	<u>3,604,070</u>	<u>(8,834,843)</u>
資產總值減流動負債	<u>6,255,204</u>	<u>(7,552,393)</u>
非流動負債		
銀行貸款	1,434,958	84,616
租賃負債	217,699	293,827
遞延收入	<u>172,991</u>	<u>180,146</u>
	<u>1,825,648</u>	<u>558,589</u>
資產/(負債)淨額	<u><u>4,429,556</u></u>	<u><u>(8,110,982)</u></u>
資本及儲備		
股本	44,584	40,281
儲備	<u>4,372,464</u>	<u>(8,132,783)</u>
本公司權益股東應佔		
權益/(權益－虧絀)總額	4,417,048	(8,092,502)
非控股權益	<u>12,508</u>	<u>(18,480)</u>
權益/(權益－虧絀)總額	<u><u>4,429,556</u></u>	<u><u>(8,110,982)</u></u>

1 編製基準

北京智譜華章科技股份有限公司(前稱「北京智譜華章科技股份有限公司Knowledge Atlas Technology Joint Stock Company Limited」)於2019年6月11日在中華人民共和國(「中國」)成立為有限責任公司。於2025年3月26日，本公司由有限責任公司改制為股份有限公司。本公司股份於2026年1月8日於香港聯合交易所有限公司(「聯交所」)主板上市。

本公司及其附屬公司(統稱「本集團」)主要從事提供大模型相關服務。

本中期財務報表乃根據聯交所證券上市規則的適用披露規定編製，包括遵守國際會計準則理事會(「國際會計準則理事會」)頒佈的國際會計準則(「國際會計準則」)第34號中期財務報告。本中期財務報表獲准於2026年8月31日刊發。

除預期將於2026年年度財務報表反映的會計政策變動外，本中期財務報表乃根據2025年年度財務報表採納之相同會計政策編製。會計政策變動詳情載於附註2。

編製符合國際會計準則第34號的中期財務報表時，管理層須作出會影響政策應用以及年初至今資產、負債、收入及開支的呈報金額的判斷、估計及假設。實際結果可能與該等估計有所不同。

本中期財務報表載有簡明綜合財務報表及選定解釋性附註。該等附註包括對了解本集團自2025年年度財務報表以來的財務狀況及表現變動屬重大的事件及交易的解釋。簡明綜合中期財務報表及隨附附註並不包括根據國際財務報告準則會計準則編製完整財務報表所需的所有資料。

2 會計政策變動

國際會計準則理事會已頒佈多項於本會計期間首次生效的新訂及經修訂國際財務報告準則會計準則。其中，僅國際財務報告準則第9號金融工具及國際財務報告準則第7號金融工具：披露(修訂本)－金融工具分類與計量的修訂與本集團的財務報表相關。應用該等修訂並未對本集團呈列期間的綜合財務報表產生重大影響。

本集團並無採用任何於本會計期間尚未生效之新訂準則或詮釋。

該等修訂亦將影響本集團截至2026年12月31日止期間之年度財務報表中將予納入的披露，內容有關指定為按公允價值計入其他全面收益的權益工具投資以及具有特定或有特徵且並非按公允價值計入損益的金融工具。本中期財務報表中並未納入額外披露。

3 收入及分部報告

(a) 收入

本集團的主要業務為提供大模型相關服務。

(i) 收入的劃分

按主要服務類型及收入確認時間劃分的客戶合約收入如下：

	截至6月30日止六個月	
	2026年 人民幣千元	2025年 人民幣千元
國際財務報告準則第15號範圍內的客戶合約收入		
按主要服務類型劃分：		
－本地化部署	128,716	161,777
－雲端部署	825,176	29,100
	<u>953,892</u>	<u>190,877</u>
收入確認時間		
某一時間點	854,270	155,636
隨時間	99,622	35,241
	<u>953,892</u>	<u>190,877</u>

(b) 分部報告

本集團按服務類型管理其業務。本集團已呈列以下兩個報告分部，其方式與本集團向最高行政管理人員內部報告資料以進行資源分配及表現評估的方式一致。概無將任何經營分部合併以形成以下報告分部。

- 本地化部署：該分部根據客戶的具體指示及需求，在客戶的基礎設施上開發並提供定製化大模型相關服務。
- 雲端部署：該分部開發並透過雲端基礎設施向客戶提供雲端大模型相關服務。

(i) 分部業績

就評估分部表現及分配資源而言，本集團最高級行政管理人員監控各報告分部應佔的業績，基準如下：

收入及開支乃參考該等分部產生的收入及該等分部產生的直接開支而分配予報告分部。報告分部業績所使用的計量方式為毛利。期內並無發生分部間銷售。一個分部向另一個分部提供的協助(包括共享資產及技術知識)不予計量。

本集團的其他經營收益及開支(例如其他收益、銷售及營銷開支、一般及行政開支、研發開支、金融資產減值虧損、融資成本以及資產及負債)不會根據個別分部計量。因此，未呈列分部資產及負債相關資料以及與資本開支、利息收益及利息開支有關的資料。

與於期內提供予本集團最高行政管理人員以進行資源分配及分部表現評估的本集團報告分部有關的資料如下。

	截至2026年6月30日止六個月		
	本地化部署	雲端部署	總計
	人民幣千元	人民幣千元	人民幣千元
來自外部客戶的分部收入	<u>128,716</u>	<u>825,176</u>	<u>953,892</u>
分部毛利	<u>48,740</u>	<u>202,871</u>	<u>251,611</u>

	截至2025年6月30日止六個月		
	本地化部署	雲端部署	總計
	人民幣千元	人民幣千元	人民幣千元
來自外部客戶的分部收入	<u>161,777</u>	<u>29,100</u>	<u>190,877</u>
分部毛利	<u>95,540</u>	<u>(116)</u>	<u>95,424</u>

(ii) 報告分部損益的對賬

	截至6月30日止六個月	
	2026年	2025年
	人民幣千元	人民幣千元
報告分部毛利	251,611	95,424
其他收益	41,436	4,614
銷售及營銷開支	(177,666)	(208,570)
一般及行政開支	(103,346)	(185,165)
研發開支	(2,131,171)	(1,594,661)
金融資產減值虧損	(27,416)	(10,867)
融資成本	(95,925)	(53,270)
分佔聯營公司利潤減虧損	144,152	14,147
按公允價值計入損益的金融工具的 公允價值變動	50,385	9,791
向投資者發行的金融工具的賬面值變動	<u>(22,090)</u>	<u>(429,295)</u>
綜合除稅前虧損	<u>(2,070,030)</u>	<u>(2,357,852)</u>

4 綜合損益及其他全面收益表的所得稅

綜合損益表的稅項指：

	截至6月30日止六個月	
	2026年 人民幣千元	2025年 人民幣千元
即期稅項	1,962	—
遞延稅項	—	—
	<u>1,962</u>	<u>—</u>

於報告期內，本集團在中國內地成立的實體須按中國企業所得稅率25%繳稅(截至2025年6月30日止六個月：25%)。

在其他司法管轄區註冊成立的附屬公司的稅收按相關司法管轄區的適用所得稅率計算。

本集團的若干附屬公司獲得了稅務機關頒發的「高新技術企業」資格，並於截至2026年及2025年6月30日止六個月均享受15%的優惠稅率。

根據中國企業所得稅法及其相關法規，符合小型微利企業資格的實體可享有5%的優惠所得稅率(應課稅收入低於人民幣3,000,000元)。本集團若干附屬公司符合小型微利企業資格，於截至2026年及2025年6月30日止六個月均享受優惠所得稅率。

5 每股虧損

每股基本虧損乃按本公司普通股權益股東應佔虧損除以報告期內已發行或被視為已發行普通股加權平均數計算。

本公司於2025年3月26日改制為股份有限公司。因此，本公司實收資本人民幣36,224,375元轉換為36,224,375股每股面值人民幣1.00元的股份。就計算每股基本及攤薄虧損而言，本公司改制為股份公司之前被視為已發行普通股的加權平均數乃假設改制為股份公司這一事件自2025年1月1日起已經發生，並按2025年3月改制時所釐定的換股比率計算。

此外，報告期內的加權平均股數亦已追溯調整，以反映本公司股份於2026年1月8日完成上市前生效之股份拆細所產生的影響。

	截至6月30日止六個月	
	2026年	2025年
本公司普通股權益股東應佔期內虧損 (人民幣千元)(附註5(a))	(2,020,788)	(993,276)
視作已發行普通股的加權平均數(千股)(附註5(b))	<u>432,359</u>	<u>159,522</u>
每股基本虧損(人民幣元)	<u>(4.67)</u>	<u>(6.23)</u>

(a) 本公司普通股權益股東應佔期內虧損

	截至6月30日止六個月	
	2026年	2025年
	人民幣千元	人民幣千元
本公司全體權益股東應佔期內虧損	(2,071,326)	(2,351,173)
向投資者發行金融工具應佔期內虧損分配	<u>50,538</u>	<u>1,357,897</u>
本公司普通股權益股東應佔期內虧損	<u>(2,020,788)</u>	<u>(993,276)</u>

(b) 普通股的加權平均數

	截至6月30日止六個月	
	2026年	2025年
	股份數目	股份數目
	千股	千股
於1月1日已發行的普通股	40,281	36,224
首次公開發售時發行股份的影響	40,361	—
視作已發行普通股的影響	—	1,536
向投資者發行金融工具的影響	(1,081)	(21,808)
股份拆細的影響	<u>352,798</u>	<u>143,570</u>
於6月30日普通股的加權平均數	<u>432,359</u>	<u>159,522</u>

(c) 每股攤薄虧損

向投資者發行的金融工具並無計入每股攤薄虧損的計算中，原因是計入該等金融工具會產生反攤薄影響。因此，截至2026年及2025年6月30日止六個月的每股攤薄虧損與每股基本虧損相同。

6 物業、廠房及設備

收購自有資產

2026年5月，本公司以人民幣360,226,000元的總代價，完成收購北京紅鑽的全部股本，包括(1)所收購股權的現金代價人民幣79,158,000元(附註9)；及(2)用於促成北京紅鑽償還相關債務的款項合計人民幣281,068,000元。於2026年8月21日，相關代價已悉數支付。

本公司選擇應用國際財務報告準則第3號企業合併所允許的集中度測試，並得出結論，收購北京紅鑽乃是資產收購而非企業合併。於交割日自北京紅鑽收購的可識別資產及負債如下：

	於交割日期 的公允價值 人民幣千元
收購的可識別資產及負債：	
物業、廠房及設備	360,000
預付款項及其他應收款項	869
所取得的現金及現金等價物	49
貿易及其他應付款項	(692)
已付/應付總代價	<u>360,226</u>

7 貿易應收款項及應收票據

	2026年 6月30日 人民幣千元	2025年 12月31日 人民幣千元
貿易應收款項	838,139	339,198
應收票據	<u>6,445</u>	<u>—</u>
	844,584	339,198
減：虧損撥備	<u>(61,640)</u>	<u>(35,990)</u>
	<u>782,944</u>	<u>303,208</u>

於報告期末，基於發票日期的貿易應收款項及應收票據(扣除虧損撥備)賬齡分析如下：

	2026年 6月30日 人民幣千元	2025年 12月31日 人民幣千元
3個月內	626,201	243,484
3個月至6個月	23,697	19,831
6個月至1年	109,486	31,406
1年至2年	23,538	8,415
2年至3年	22	72
	<u>782,944</u>	<u>303,208</u>

本集團一般要求本地化部署及相關服務的客戶結算進度款項，而雲端部署客戶則須根據按用量計費及相關服務合約預先付款或自發票日期起計0至60天內支付，並根據按訂閱模式計費的合約及相關服務進行定期結算。

8 預付款項及其他應收款項

	2026年 6月30日 人民幣千元	2025年 12月31日 人民幣千元
存款	78,404	70,538
出售按公允價值計入損益的股本證券投資 所產生的應收款項	—	7,098
其他應收款項	<u>110,379</u>	<u>52,029</u>
	188,783	129,665
減：虧損撥備	<u>(20,923)</u>	<u>(18,994)</u>
	<u>167,860</u>	<u>110,671</u>
預付計算服務費及其他款項	1,243,736	116,295
可抵扣的進項增值稅	<u>209,715</u>	<u>168,768</u>
	<u>1,453,451</u>	<u>285,063</u>
	<u>1,621,311</u>	<u>395,734</u>

9 貿易及其他應付款項

	2026年 6月30日 人民幣千元	2025年 12月31日 人民幣千元
貿易應付款項	241,589	297,139
計算服務費應付款項	1,427,415	727,348
營銷及推廣服務應付款項	51,004	42,812
應付僱員成本	122,923	146,507
企業合併及資產收購應付代價(附註6和11)	229,809	—
其他應付款項及應計費用	26,869	54,437
按攤銷成本計量的金融負債	2,099,609	1,268,243
其他應付稅項	17,515	39,774
保修撥備	18,635	20,614
	2,135,759	1,328,631

於報告期末，基於發票日期的貿易應付款項賬齡分析如下：

	2026年 6月30日 人民幣千元	2025年 12月31日 人民幣千元
3個月內	171,575	229,890
3個月至6個月	15,254	46,448
6個月至1年	50,208	19,506
1年以上	4,552	1,295
	241,589	297,139

10 股息

本公司董事不建議就截至2026年6月30日止六個月派付中期股息(截至2025年6月30日止六個月：無)。

11 企業合併

於2026年6月5日，本集團通過本公司與北京中科加禾智能科技有限公司(「**中科加禾**」)的權益持有人訂立收購協議，以人民幣291,916,000元的總現金代價收購中科加禾60%的股權。根據收購協議，代價將分四期支付，且人民幣141,264,000元已於2026年6月支付。中科加禾主要從事AI基礎設施研發。該收購事項已於2026年6月完成。

從中科加禾收購的可識別資產及負債的公允價值載列如下：

	於收購日期 人民幣千元
可識別資產及負債：	
物業、廠房及設備	1,309
無形資產	22,510
遞延稅項資產	3,377
存貨	3,934
貿易應收款項及應收票據	1,373
預付款項及其他應收款項	209
銀行存款及手頭現金	66,567
貿易及其他應付款項	(6,558)
合約負債	(9,906)
租賃負債	(304)
遞延稅項負債	(3,377)
已收購的可識別資產淨值的公允價值	79,134
減：非控股權益	(31,654)
加：商譽	244,436
收購總代價	<u>291,916</u>

收購中科加禾於收購日的現金及現金等價物現金流出淨額分析：

人民幣千元

現金代價總額	291,916
減：收購日所取得的現金	(66,567)
減：應付代價(附註9)	<u>(150,651)</u>
計入簡明綜合現金流量表的收購一間附屬公司的 現金流出淨額	<u><u>74,698</u></u>

12 報告期後非調整事項

於2026年7月13日，本公司完成一項股份配售，以每股1,588港元之價格發行19,780,000股普通股。配售所得款項總額約為31,410,640,000港元(相當於約人民幣27,230,826,000元)。

其他資料

中期股息

董事會不建議派付截至2026年6月30日止六個月的中期股息(截至2025年6月30日止六個月：無)。

購回、出售或贖回本公司上市證券或出售庫存股份

購回授權

本公司董事會已根據股東於2026年6月22日通過的決議案獲授予一般授權(「購回授權」)，以不時於公開市場上購回股份。根據購回授權，本公司獲准購回不超過2025年年度股東會通過前述議案之日本公司已發行H股股份總數10%的股份。

本公司H股於2026年1月8日首次在香港聯交所主板上市。自上市日期起及直至2026年6月30日止期間，本公司或其任何附屬公司概無購回、出售或贖回本公司於聯交所上市的任何證券(包括出售任何庫存股份)。

於2026年6月30日，本公司並無持有上市規則所定義的任何庫存股份。

報告期後重大事項

根據一般授權配售新H股

於2026年7月9日(交易時段前)，本公司與中國國際金融香港證券有限公司(「獨家配售代理」)訂立配售協議(「2026年7月配售事項」)。據此，本公司同意委任獨家配售代理，而獨家配售代理同意作為本公司之代理盡最大努力促使不少於六(6)名承配人(將為專業、機構或其他投資者，且彼等及彼等之最終實益擁有人將為獨立第三方)根據本公司於2026年6月22日舉行之年度股東會授予董事之一般授權，按配售價每股配售股份1,588.00港元認購最多19,780,000股新H股(「配售股份」)。2026年7月配售事項已於2026年7月13日完成。2026年7月配售事項所得款項淨額(經扣除佣金及估計開支後)合共約為31,374.95百萬港元。詳情請見本公司日期為2026年7月9日及2026年7月13日的公告。

除上文及本公告所披露者外，自報告期結束後至本公告日期，並無其他可能影響本公司的重大事項。

遵守企業管治守則

本公司明白達致高水平企業管治標準的價值及重要性，以提升企業表現、透明度及問責性，並進一步贏得股東及公眾的信任。本公司致力維持高水平的企業管治，並相信此對本公司的發展及保障股東的權益至為重要。企業管治乃董事會指示本集團管理層處理其事務之程序，以確保達致其目標。

自上市日期起至本公告日期，本公司嚴格遵循企業管治守則的所有適用守則條文。此外，本公司會積極參考企業管治守則的建議最佳常規，包括建立完善的內部控制體系、強化信息披露的即時性與準確性，保障股東知情權與參與權，持續優化治理結構。

遵守標準守則

本公司自上市以來已採納標準守則作為規管其董事及可能獲得有關本集團或本公司證券內幕消息的僱員進行本公司證券交易的行為守則。

經向全體董事作出具體查詢後，全體董事均已確認彼等自上市日期起及直至2026年6月30日止期間一直遵守標準守則。

可能擁有本集團內幕消息的本集團僱員須遵守標準守則。自上市日期起及直至2026年6月30日止期間，本公司並無發現相關僱員違反標準守則的任何事件。

上市規則規定的持續披露責任

除本公告所披露者外，本公司概無根據上市規則第13.20、13.21及13.22條規定的任何其他披露責任。

公眾持股量之充足性

根據公開披露資料及董事所悉知，自上市日期起及直至本公告日期，本公司一直維持上市規則規定的公眾持股量。

審計委員會

審計委員會由謝德仁博士、楊強博士及李涓子博士組成。謝德仁博士為審計委員會主席。審計委員會已審閱本集團截至2026年6月30日止六個月的中期業績，及按國際財務報告準則會計準則編製的本集團截至2026年6月30日止六個月的未經審計簡明綜合財務報表。董事會與審計委員會之間並無就本公司於報告期所採納的會計處理方法出現分歧。

中期財務資料的審閱

本中期業績公告所載財務資料摘錄自截至2026年6月30日止六個月的中期財務報告，該中期財務報告未經審計，但已由畢馬威會計師事務所根據香港會計師公會頒佈的香港審閱工作準則第2410號「由實體的獨立核數師執行中期財務資料審閱」進行審閱。畢馬威會計師事務所已出具無修改結論的審閱報告，該報告將載於寄發予股東的中期報告內。

刊發中期業績及中期報告

本業績公告已刊發於聯交所網站(www.hkexnews.hk)及本公司網站(www.zhipuai.cn)。本公司截至2026年6月30日止六個月的中期報告將適時刊發於上述聯交所及本公司網站，並按本公司H股股東選擇接收公司通訊的方式寄發予本公司H股股東。

釋義及技術詞彙表

於本公告內，除文義另有所指外，下列詞彙具有下文所載涵義：

釋義

「審計委員會」	指 董事會審計委員會
「董事會」	指 本公司董事會
「企業管治守則」	指 上市規則附錄C1所載《企業管治守則》
「中國」	指 中華人民共和國，僅就本公告而言，不包括香港、澳門特別行政區及台灣
「本公司」或「公司」	指 北京智譜華章科技股份有限公司，一家於2019年6月11日根據中國法律成立的有限責任公司，並於2025年3月26日改制為股份有限公司，其H股在聯交所上市(股份代號：2513)
「董事」	指 本公司董事

「全球發售」	指	招股章程所界定的香港公開發售及國際發售
「本集團」或「我們」	指	本公司及其附屬公司
「H股」	指	本公司每股面值人民幣0.10元的境外上市股份，以港元買賣並於聯交所上市及買賣
「港元」	指	港元，香港法定貨幣
「香港」	指	中國香港特別行政區
「國際財務報告準則會計準則」	指	國際會計準則理事會頒佈的國際財務報告準則會計準則
「獨立第三方」	指	獨立於本公司及其關連人士且與彼等概無關連的第三方
「上市」	指	H股於2026年1月8日於聯交所主板上市
「上市日期」	指	2026年1月8日，H股在聯交所主板上市的日期
「上市規則」	指	《香港聯合交易所有限公司證券上市規則》
「標準守則」	指	上市規則附錄C3所載《上市發行人董事進行證券交易的標準守則》
「招股章程」	指	本公司日期為2025年12月30日內容有關全球發售及上市的招股章程
「人民幣」	指	人民幣，中國法定貨幣
「報告期」	指	截至2026年6月30日止六個月
「股份」	指	本公司股本中每股面值人民幣0.10元的普通股，包括未上市股份以及H股
「股東」	指	股份持有人

「聯交所」或「香港聯交所」	指	香港聯合交易所有限公司
「附屬公司」	指	具有上市規則賦予該詞的涵義
「庫存股份」	指	具有上市規則賦予該詞的涵義
「未上市股份」	指	本公司發行的普通股，每股面值人民幣0.10元，未在任何證券交易所上市
「美元」	指	美元，美國法定貨幣
「%」	指	百分比

技術詞彙表

詞彙

定義

「Agent/智能體」	指	能夠自主執行多步任務鏈的AI模型，可在無人逐步指示的情況下自行調用工具、處理中間失敗並完成多步任務
「Agentic Engineering」	指	模型從簡單代碼生成走向完整工程項目交付的能力階段，模型需持續理解目標、調用工具、處理異常、並根據環境反饋調整策略，以實現端到端的工程交付
「AGI商業價值公式」	指	本集團對AGI商業價值的量化框架，公式為： $AGI商業價值 \Rightarrow (任務價值密度 \times Token消耗規模)$ ，其中智能上界決定任務邊界，任務邊界決定Token所處的價值密度
「All-in-Infra戰略」	指	本集團自2026年初全面實施的基礎設施戰略，在訓練與推理兩側同時投入，並將國產芯片納入主力推理算力，以掌握自身成本曲線
「算力乘數」	指	每投入1元算力(含訓練與推理)所對應的開放平台及API收入，用以衡量模型的自我造血能力及算力投入回報效率
「Co-work」	指	模型能力階梯中的第四級，即模型交付可被專業人士覆核的工作成果，模型進入真實工作流，按完成任務計價；門檻為可靠性達到專業人士願意覆核而非推倒重做的水平

「Chat」	指	模型能力階梯中的第一級，即模型交付單次回答，任務在一輪之內結束，價值密度最低但覆蓋最廣
「Coding」	指	模型能力階梯中的第二級，即模型交付一段可運行的代碼，首次出現客觀驗收標準(能否跑通)；從Chat到Coding的門檻為結果可驗證
「CyberGym」	指	用於評估AI模型在網絡安全領域能力的基準測試平台
「ExploitGym」	指	用於評估AI模型漏洞利用能力的基準測試平台
「Scaling deep」	指	深入後訓練、強化學習、長程軌跡與任務反饋的Scaling路徑，與Scaling up(擴大模型規模、上下文及算力)相對
「Scaling up」	指	通過擴大模型規模、上下文長度、算力與系統承載來提升模型能力的路徑
「GLM-5.3-Flash(Ox-Alpha)」	指	本集團發佈的下一代輕量化旗艦模型，測試代號Ox-Alpha，採用全新架構，總參數320B、激活參數18B、層數45層，採用稀疏注意力與線性注意力混合架構，以極低成本實現高性能推理，設計目標為成本前沿
「GLM-5.3」	指	本集團發佈的旗艦大模型，與GLM-5.2同架構、同總參數、同激活參數，唯一變量為後訓練規模，端到端完成率較GLM-5.2提升超過50%，代表當前模型能力上界

「中訓練」	指	本集團在預訓練與後訓練之間單獨配置算力的訓練階段，用於提升知識密度、擴展有效上下文長度、建立工具調用與長程任務的基礎模式
「Critic-based SAO」	指	本集團自研的基於評估器的單軌跡優化算法(Critic-based Single-trajectory Advantage Optimization)，由GRPO探索演化而來，針對長軌跡在獎勵分配、反饋利用與持續學習方面的挑戰
「OPD」	指	本集團在算法維度由監督微調(SFT)延伸探索的訓練方式，全稱為離線策略蒸餾(Offline Policy Distillation)
「Index Share」	指	本集團自研的技術，用於降低長上下文稀疏注意力開銷
「MTP」	指	多Token預測(Multi-Token Prediction)技術，用於優化推理解碼效率
「LayerSplit」	指	本集團自研的技術，用於優化KV Cache管理與集群網絡效率
「ZCube」	指	本集團自研的技術，與LayerSplit協同用於優化KV Cache管理與集群網絡效率
「mHC(流形約束超連接)」	指	GLM-5.3-Flash所採用的架構創新，全稱為流形約束超連接(manifold-constrained Hyperconnection)，用於提升模型的Scaling能力
「Encode-Prefill-Decode分離式架構」	指	本集團在生產級集群層面採用的推理架構，將多模態編碼、prompt預填充與逐Token解碼拆成可獨立調度、獨立擴縮容的工作池，以優化國產芯片上的推理性能

「Loop Transformer (循環架構)」	指	通過層的循環復用，在不顯著增加參數量的前提下提高有效計算深度的模型架構，將算力從「記」轉向「想」，適合需要長程推理的高複雜度任務
「Model-vs-Model自我對弈管線」	指	本集團自研的合成數據生成機制：由一個模型生成解法與內容，另一個模型評審、測試、執行驗證，通過對抗檢驗的數據才進入訓練環節，並經四層過濾(規則校驗、執行驗證、模型評審、人工抽檢)後用於預訓練、中訓練與後訓練
「Fully Self Training」	指	本集團的技術願景，即下一代GLM模型將在上一代GLM所搭建的環境裡訓練，形成遞歸式自我改進(Recursive Self-Improvement)閉環，涵蓋數據自產、環境自造、基礎設施自我優化三個維度
「遞歸式自我改進 (Recursive Self-Improvement)」	指	模型參與自身改進所形成的遞歸閉環，包括由模型生成環境用於訓練下一代模型、以及由模型優化服務自身的推理引擎
「安全披露賬本 (CVD)」	指	本集團建立的公開安全披露機制，記錄漏洞從發現、確認到修復的全過程，僅公開已完成披露並由廠商修復的漏洞；對仍處於協調披露階段的漏洞，公開哈希值以供後續核驗
「OpenVuln」	指	本集團「開源的盾」計劃下公開的安全審計結果平台，用於向開源社區披露對重點開源項目的持續安全審計成果

「ZCode」	指	本集團推出的編程智能體產品，通過Harness承擔上下文管理、工具調用、任務調度、緩存與結果校驗，改善模型在真實工程環境中的表現，並提供代碼審計功能
「GLM Coding Plan」	指	本集團於2026年8月向全量用戶開放的編碼智能體訂閱服務，覆蓋Lite、Pro、Max及團隊版本，是國內最早以訂閱制交付編碼智能體能力的大模型服務之一
「能力階梯」	指	本集團對模型能力演進路徑的描述框架，依次為：Chat → Coding → Agent → Co-work → Autonomous AI，每一級向上均有具體技術門檻，且門檻跨越後方可解鎖上一層的商業模式
「Autonomous AI」	指	能力階梯中的第五級，即模型交付持續運行的能力，在無人值守條件下承擔長期目標；門檻為模型能夠自主判斷任務完成質量
「智能指數 (Intelligence Index)」	指	由Artificial Analysis編製的模型智能綜合評分，用於衡量模型的整體能力水平
「帕累托前沿 (Pareto Frontier)」	指	在成本與智能兩個維度上同時最優的模型集合，即在給定成本下智能最高、或在給定智能下成本最低的模型所構成的前沿
「任務價值密度」	指	每單位智能消耗(即每單位Token)對應的經濟產出；相同規模的Token用於不同任務(如閒聊對比修復生產環境漏洞)所產生的價值差距可達數個數量級

為便於參考，本公告以中文及英文載列中國法律法規、政府機關、機構、自然人或其他實體名稱，倘有任何歧義，概以中文版為準。

本公告載有若干前瞻性陳述。該等前瞻性陳述乃基於本集團目前可獲得的資料或董事會當前的信念、預期及假設而作出。該等前瞻性陳述受風險、不明朗因素及其他非本公司所能控制的因素影響，可能令實際結果或業績與該等前瞻性陳述所明示或暗示的情況有重大差異。鑒於存在風險及不明朗因素，本公告中包含的前瞻性陳述不應被視董事會或本公司對計劃和目標將實現的陳述，且股東及本公司投資者不應過分依賴該等陳述。

除另有指明外，本公告所載涉及第三方機構、公開榜單、評測平台或行業研究資料的所有圖表、排名及數據乃引用自公開可得的資料，僅供參考及說明用途。本公司並無參與該等數據、榜單或評測結果的編製、計算或驗證程序，亦不對其方法論、樣本範圍、準確性、完整性或時效性作出任何陳述、保證或承擔任何責任。

承董事會命
北京智譜華章科技股份有限公司
執行董事兼董事長
劉德兵博士

香港，2026年8月31日

於本公告日期，董事會成員包括：(i)執行董事劉德兵博士、張鵬博士及張笑涵女士；(ii)非執行董事李涓子博士、李家慶先生及王盟先生；及(iii)獨立非執行董事楊強博士、謝德仁博士及徐文鳴先生。