Accelerating AI Industry Evolution Reshapes Value Distribution Across the Supply Chain

AI technologies continue to evolve rapidly both in China and overseas, with underlying industry fundamentals remaining on an upward trajectory. However, global technology markets have experienced a synchronized pullback over the past week. Zhang Jingqi, General Manager of the Proprietary and Dedicated Accounts Investment Department at Rosefinch Fund, attributes this weakness partly to rising U.S. Treasury yields, growing concerns over returns on AI capital expenditure and financing models, and a broad-based deterioration in risk appetite across domestic and overseas equity markets.

Since September, OpenAI has introduced its GPT-6 Astra, Sol, and Luna model series, improving capabilities while lowering prices, further driving demand for computing power rentals. Anthropic, meanwhile, has proposed slowing model releases and strengthening pre-deployment safety reviews. While this may appear to constrain training demand, safety validation itself consumes substantial additional computing resources. Meta's launch of MUSE Agent has also made the prospects for consumer-facing AI agents increasingly tangible. These agents could automate a wide range of routine tasks, including form filling, price comparisons, and ticket booking, while generating incremental demand for CPUs.

Leading semiconductor companies have announced an additional USD 150 billion in share buybacks, bringing total buybacks to USD 235 billion. Their market capitalization has approached USD 6 trillion, with the stock reaching a new high at one point. Leading memory companies reported strong earnings growth in FY26 Q4, while customers have secured forward HBM capacity through long-term agreements. The market is increasingly recognizing memory as an asset class characterized by more stable earnings rather than purely cyclical performance. Meanwhile, leading foundries have reported strong revenue growth and plans to raise prices for advanced-node manufacturing, reinforcing the high visibility of downstream AI demand.

At the same time, structural changes are emerging within the industry. The rapid development of open-source models has expanded industry supply, while the rise of agentic AI has driven overall AI demand higher. When open-source models are used, cloud service providers generally retain greater bargaining power and a larger share of revenue, potentially generating better returns on invested capital (ROIC). This, in turn, has placed some downward pressure on the valuations of leading closed-source model developers.

Recently, however, leading overseas closed-source model developers have continued to refine and upgrade their models, gradually translating their performance advantages into economic advantages. Costs per unit of intelligence and per task continue to decline. Anthropic's pricing per token has even fallen below that of some domestic peers in China.

Jevons' paradox remains relevant: as model costs continue to fall, total AI demand is likely to expand further

rather than reduce overall computing consumption as efficiency improves. However, the mechanism governing value distribution across the supply chain is changing. Leading closed-source model developers are gaining pricing power and greater influence over the industry, squeezing the revenue share available to cloud service providers. Emerging cloud service providers are particularly exposed to these pressures.

If this trend continues, the scenario described by Anthropic CEO Dario Amodei may become increasingly plausible: value in the AI industry could become highly concentrated among leading foundational model developers, with other companies taking on more supporting roles.

This creates a practical dilemma for cloud service providers. In the short term, model costs are falling and demand remains robust, but bargaining power is shifting increasingly toward model developers, raising questions about cloud providers' return on investment (ROI). One potential solution is to bring computing capacity online as quickly as possible, before the capability gap between closed-source and open-source models widens further, thereby securing a share of the resulting revenue opportunities. However, large-scale computing infrastructure requires substantial external financing, leaving cloud providers caught between the need to invest aggressively and concerns over funding and returns.

Over the longer term, leading model developers such as OpenAI and Anthropic are likely to continue increasing the proportion of computing capacity they build and operate themselves. Elon Musk's companies, meanwhile, have rapidly deployed infrastructure by leveraging their execution capabilities and secured long-term computing rental contracts in advance, providing a reference case for the industry. Against this backdrop, demand for hardware should continue to grow. The market may increasingly favor companies capable of deploying computing infrastructure quickly and securing long-term computing contracts, assigning them higher valuations.

At the capital markets level, global technology assets have declined in tandem over the past week. Following the preceding correction, valuations and trading congestion in China's A-share technology sector have eased to some extent, while underlying fundamentals remain strong. The sector's relative investment attractiveness therefore warrants attention.

Looking ahead, major domestic and overseas technology companies will release their third-quarter earnings in a concentrated period in late October. Beyond reporting business results already achieved, their guidance on AI capital expenditure, investment returns, and the commercialization of AI applications will be particularly important, providing key signals for the next phase of the industry's development.

From an investment perspective, AI hardware remains a core area of focus. Key segments include domestic computing infrastructure, printed circuit boards (PCBs) and copper-clad laminates (CCLs), memory, and high-speed connectivity. Investors should avoid excessive concentration in any single segment. At the same time, continued attention should be paid to progress toward commercially viable business models for AI applications, particularly marginal changes in commercialization data across AI agents, office productivity, video, and related areas.

Disclaimer

We endeavor to provide timely English version of Rosefinch's publications. For any English translation of the original Chinese article, in case of any discrepancy, the Chinese version takes precedence.

The information contained herein is for reference purposes only. It does not constitute a necessary basis for future investment decisions of any funds managed by the Company, nor does it represent any substantive investment advice or commitment to readers or investors. The Company makes no representations or warranties as to the accuracy or completeness of the information and data contained herein, and shall not be held liable for any investment outcomes incurred by any third party as a result of reliance on this document. Investments involve risks, and investors should exercise caution.

The views expressed herein are those as of the date of this document and are subject to change. Rosefinch Fund Management may issue opinions at other times that are inconsistent with those expressed herein. Without prior written consent from Rosefinch Fund Management, no institution or individual may reproduce, distribute, publish, or cite this document in any form.