The latest notes on DeepSeek's all-in-one machine: the full-parameter version is not selling well, and the market is grabbing scattered low- and mid-end models

Note: These are Jeffrey Ding's informal and unofficial translations -- all credit for the original goes to the authors and the original text linked below. These are informal translations and all credit for the original work goes to the authors. Others are welcome to share ***excerpts*** *from these translations as long as my original translation is cited. Commenters should be aware that the Google Doc is also publicly shareable by link. These translations are part of the ChinAI newsletter - weekly-updated library of translations from Chinese thinkers on AI-related issues: https://chinai.substack.com/*

Authors: Erica Zhao (Zhiqi Zhao)
Source: AItechtalk, under Leiphone
Date: July 25, 2025
Original Mandarin: https://mp.weixin.qq.com/s/jQWybc7aPWOfZ5PbWTJRPw

There are no new KA (key account) customers, and no returning customers. What is the next step for the all-in-one machine market?

Two weeks ago, a certain state-owned enterprise in Guangdong issued a 10,000-word document, bidding for a DeepSeek all-in-one machine with a budget of 400,000 RMB.

In the nearly 30-page document, the purchase details table is only 6 lines long, and key indicators such as the required chip type are not mentioned at all. "Looking at this bid, you can see that no one has used the all-in-one machine at all," said an industry insider.

The needs of new users are still vague, so will old users come back? Leiphone asked many industry insiders, and the answers were all negative. Only one leading complete-system vendor had repeat sales — repeat purchases for two machines from a research institute. "Everyone is wallowing about in the pit where the first batch of users landed, and new users are more cautious," lamented Jianing, a complete-system vendor sales representative.

Four months ago, all-in-one machines were the hottest favorites in the market. Related words were everywhere in industry conferences, technology exhibitions, and airport advertisements.

Four months later, it has become a thing of the past. In the product catalogs of many manufacturers, all-in-one computers occupy a place, but there are not many sales, even cloud giants may not be immune. Some R&D personnel who work on all-in-one machines in large tech companies can't help but worry: "If this continues, should I find a new job?"

But looking at it from another perspective, the situation is completely different-

Tens of billions of market revenue and thousands of monthly sales have been easy for leading hardware manufacturers in the past six months. After all the hype, everyone realized that the real moat of the all-in-one machine business is not technology, but the customer relationships that come from (companies with) hardware genes.

So, with rich customer relationships, can the leading manufacturers still have such a big cake in the second half of the year?

In this regard, the new consensus reached by everyone is: the market for high-end models (mainly deployed with full-parameter versions) has basically ended, and what remains is the guerrilla warfare of small and medium-sized markets or low- and mid-end models.

Who will join this guerrilla war? What preparations have everyone made? Do the new entrants who are ready to fight still have a chance? And what is the current situation of the all-in-one machines that have been sold?

01. First half of the year: The top hardware manufacturers are crowded with customers, while the others have unfulfilled ambitions

In February, the domestic market, which has not seen a new growth point for ToB for a long time, had high hopes for LLM all-in-one machines.

At that time, SOE’s had a strong demand for high-end models, and the revenue of some operators reached at least 1 billion. The market scale of some leading all-in-one machine vendors in the first half of the year was close to 10 billion, and some hardware manufacturers could still achieve a monthly sales volume of at least 1,000 units of various models in May.

"In short supply" is the summary of the sales of all-in-one machines in the first half of the year by an insider at IEIT Systems (Inspur subsidiary). In the past six months, when the market supply was somewhat tight, "buy whoever has the goods" once became an important logic for users to choose manufacturers.

Therefore, the price of all-in-one machines rose sharply. Industry insiders have observed that after DeepSeek went viral, H20 "has a new price per week", and each price increase is about 10%. The price of an all-in-one machine equipped with an Nvidia H20 card from a leading manufacturer has risen from 1.3 million to around 1.5 million in two months. But the industry insider also added that around April, only 10-20% of the transactions were sold at around 1.5 million, and more were sold at 1.3 million.

Amid price fluctuations, the benefits of this trend to hardware manufacturers can be seen from their financial reports and revenue. IEIT Systems's financial report shows that its net profit in the first quarter of 2025 was about 460 million RMB, and its revenue increased by 165.31% year-on-year. The significant growth in server sales is the core driving force for performance. H3C's financial report also shows that the company's revenue in Q1 increased by 26.07% year-on-year, of which domestic government business revenue was 12.255 billion RMB, a year-on-year increase of 37.49%.

Hardware manufacturers are engaged in heated battle, while small and medium-sized manufacturers have found scanty business . The number of orders from many manufacturers is only in the low double digits, mainly deploying 32B or 70B models; or just listing the all-in-one machine as a category in the product list, and not promoting it in sales strategy. There are also independent cloud vendors that hope to urgently pull up the upstream and downstream to provide products after customers place orders - however, without getting the early parameters, users cannot make up their minds to place orders. Not to mention the channel dealers who enter the market halfway, many of them end up with the goods in their own hands.

One of the important reasons for these "two worlds" of sales is the customer relationship of the enterprise itself. The key groups in the deployment boom in the first half of the year, such as state-owned enterprises, financial institutions and other large customers, often have their own hardware procurement lists. The companies on the list generally have to meet strict supplier qualifications and business conditions. In other words, they are all integrated in fixed sales channels in the past.

In this field, the old *guanxi* connections of the leading hardware manufacturers will undoubtedly be stronger. The crucial moat in the all-in-one machine business that the market struggled to figure out in the first half of the year turned out to be customer relationships.

In another aspect, customers whose deployment needs are driven by strategy only want to solve the problem of "having or not having" an all-in-one machine, and do not pay much attention to the software layer. It is reported that even for the leading hardware manufacturers, the average shipment rate of supporting software is only 10% to 20%. However, the platform capabilities of the complete-system vendors vary: some manufacturers simply combine the data operation function and the large model function into one; for others, users even need to pay licensing fees to two different departments when purchasing this set of products.

In addition, there is not much room for software companies with stronger upper-level platform capabilities in this wave of business. Although the aforementioned large customer types are accustomed to purchasing software and hardware separately, at this stage, without good enough applications, the domestic market, which already lacks payment consciousness, has no motivation to specifically ask software firms to deploy. According to statistics from a software firm, the consulting opportunity conversion rate of their all-in-one (machine) business is only about 5%.

In large tech companies, the all-in-one business is often placed in the cloud business department, but some sales people in cloud giants lament that many users only want to buy hardware, and the all-in-one machines with high prices after binding software are not very popular. In addition, the actual implementation of the all-in-one machine is not as "out-of-the-box" as promoted in advertisements. For private deployment, vendors need to incorporate user data for on-site debugging and educate customers. For cloud vendors who prefer "volume (easily scalable)" business and full-platform business, it is not a sexy business.

However, the attitudes of large tech giants towards all-in-one machines are also different. Some are mainly sold by partners, while others have not clearly launched all-in-one machine-related businesses; some have launched products and have leaned on sales, but are suffering from the limitation of real landing scenarios, and they also say frankly that they are "still exploring".

In the first half of the year, the all-in-one machine market was driven by deployment demand strategy, and users pursued KPI psychology significantly. Buying hardware and shoveling away software is quite like "buying the wooden box and returning the pearls".

02. Second half of the year: "Guerrilla War" of low- and mid-end all-in-one machines begins

As an all-in-one machine salesperson, Yang Fei has finally run the all-in-one machine business into a more "standardized" model in the past three months-at least, when he lands the equipment for customers, the time he needs for reliability testing has been shortened from more than one month at the beginning to only one or two weeks now.

However, just as he was becoming familiar with this business line, the market enthusiasm gradually cooled. People like him who are fighting on the front line of the market now have a common feeling about the market: the demand for large-scale deployment has ended.

All this happened gradually. At the beginning of March, everyone was still complaining about the low conversion rate of business opportunities, but since April, even the business opportunities of "only consulting, no orders" have dropped from hundreds of opportunities a day to more than ten. In various WeChat groups, it became rare to see users asking and discussing all-in-one

machines in all places. Customers' budgets are also tightening, often only around 200,000 to 400,000 RMB, and a single company generally purchases 1-2 units.

After that, the budget scale of the bidding for all-in-one machines on the market has also been reduced from tens of millions to millions, and now it is mainly hundreds of thousands. Even if there is a bid of 300,000 or 400,000, the top candidate vendor might also reduce the price to about 200,000 RMB. These demands are more from colleges and universities, local enterprises and municipal governments in third-, fourth- and fifth-tier cities, and are basically no longer seen in first- and second-tier cities.

In terms of product requirements, customers no longer blindly pursue full-parameter versions, and will actively propose to deploy 70B and below models, and also bring actual business needs. For this reason, Yang Fei will also help connect independent software vendors and provide more industry-specific solutions. Talking about his role, he sighed: "It's just about earning hard-earned money."

Yang Fei also knows that the all-in-one machine business is always phased. However, the reality is that even if it is in the top hardware manufacturers, can the company still bear the high costs brought by the early high-end full-parameter version of equipment?

Even Inspur Electronic Information Industry (IEIT), which had a large-scale increase in revenue in the first half of the year, mentioned in its Q1 financial report that the company's short-term loans increased by 279.55%, due to the increase in capital demand for business expansion; and the asset impairment loss increased by 63.5% year-on-year. According to analysis, it was caused by the increase in inventory scale and the fluctuation of raw material prices. The impact of costs on leading vendors should not be underestimated.

Due to the need for full-blooded performance, there were very few chip manufacturers available for high-end models at that time. But after entering the non-full-blood version market, the model's configuration requirements are not so stringent. When purchasing chips, everyone has begun to look for Chinese chips with higher cost-effectiveness.

For mid- and low-end models, the profit improvement brought by chip changes cannot be underestimated: as long as the cost of each chip is reduced by two or three thousand, the cost of a four-card machine may be reduced by more than ten thousand. Some analysts have also run the numbers: when a software vendor chooses a chip from a leading domestic chip manufacturer, the net profit margin is only in the single digit; but if it chooses Hygon's chips, its net profit margin may double.

Chinese chips are also beneficiaries of the all-in-one machine wave — if you add in the small manufacturers who had customers with IT localization needs in the early days that were unable to get the best chips, so their demands were dispersed to more domestic chip manufacturers. This year, the shipment volume of domestic chips will see an increase.

For the whole year, some leading complete-system vendors said that this year, they will sell at least thousands of all-in-one machines run on high-end Huawei Ascend 910B cards — this number may have been only in the double digits last year; combined with the sales targets of mid- and low-end models, their hardware revenue is expected to triple to five times that of last year. Some industry insiders believe that it is not difficult for these manufacturers to achieve such a goal.

But the all-in-one machine market is also a besieged city. After Leiphone released the article "Regarding the truth of DeepSeek all-in-one machine landing, we surveyed 12 listed companies" in April this year, many vendors were eager to enter the all-in-one machine business. However, as the market enthusiasm declined, when these companies were contacted again some time ago, all of them were not interested in it, and frankly admitted that this business line had been shelved and not promoted.

However, there are still a few vendors who are ready to launch large-model all-in-one machine business in the second half of the year. Some will be launched in combination with self-developed chips; some will be shipped soon and set an initial revenue target of more than 200 million (RMB).

It is worth mentioning that these vendors who still have the courage to enter the market at this time have one thing in common: they all have good government relations, and even if the products have not yet landed, the orders are already in hand.

03. Have the all-in-one machines that have been sold been used?

Yet, DeepSeek all-in-one machines have almost only early adopters, but no repeat customers. What went wrong?

Previously, Leiphone has analyzed in detail the challenges faced by users in the early stage of the market outbreak in its article about the truth of DeepSeek all-in-one machine landing. These include: the choice of purchase channels, illusions about the emergence of knowledge bases, the strategic game between the full-parameter version and distilled versions, etc.

In the past few months, users are also becoming disenchanted with all-in-one machines.

On the one hand, there is a gap between the non-full-parameter version model and the user's imagination. Jianing observed that in the actual landing, the deployment of RAG (retrieval-augmented generation) and Agent may be able to make a small demo with 70B, but if you want to run stably and smoothly, it is a bit powerless.

In addition, in the overwhelming and bustling news of "xx all-in-one machine and xx application have completed adaptation" some time ago, there is also a lot of watered-down content. This is a common trial-and-error strategy in the industry - after reaching a deal, advertise first, provide users with enough options from application manufacturers, and then truly adapt, customize development and deploy according to user needs. Some research institutions have found that there is generally a month's gap between the public announcement and actual adaptation of the vendor's all-in-one machine.

On the other hand, there is a problem that must be considered after the market calms down: ROI. In the first half of the year, state-owned enterprises, governments and financial institutions that were just trying to complete the deployment task didn't grasp the account's ROI. The same is true for users who are eager to try new things. Some investment fund companies spent hundreds of thousands just to buy a 70B machine to experience it and "catch up on the fun."

As mentioned earlier, most customers now have a budget of around 200,000 to 400,000 RMB. In addition to the fact that many corporate IT budgets were planned at the end of last year, this also reflects everyone's distrust of the value of AI landing - it has not yet truly empowered the business.

The main deployment scenarios of DeepSeek all-in-one machines today are still knowledge bases, office applications, and Q&A. However, is it cost-effective to spend hundreds of thousands or even millions to deploy office applications? Many decision makers have already included this issue in their thinking. Moreover, the all-in-one machine is OK for writing structured files such as government documents, but it seems to be difficult for more flexible files.

Even in the seemingly broad application scenarios of hospitals, the implementation of all-in-one machines faces certain levels of awkwardness: (all-in-one machines deployed in) large hospitals may not be able to support enough concurrent users; and outmoded/backward informatization in small hospitals may become a barrier to the entry of all-in-one machines. The data island problem between different departments and hospitals is serious, and data quality needs to be improved.

Some all-in-one machine vendors have previously contacted users in several hospitals, but they feel that "hospital leaders do not understand AI so deeply". They are often called to meetings and no questions are asked after the demonstration, and there is no follow-up.

However, there are also real implementation cases of all-in-one machines, but this type of implementation requires more specific industry models and local deployment combined with user data.

The growth of this type of implementation demand is not as vigorous as that of hardware manufacturers, and is often more gradual. For example, a vendor in the insurance field expects to sell 500 machines this year. At present, it is steadily advancing, and "some orders will be received one after another." A certain AI company that is good at computer vision algorithms

has also made its all-in-one machine business generate tens of millions of RMB in revenue this year due to its advantages in the vision field.

A server dealer observed that among the 20 or 30 all-in-one machines he sold in the first half of the year, the one that was actually used was a foreign trading company that bought an H20 and deployed their own model to provide intelligent customer service.

The reason why this customer's cost-benefit can be calculated is that the store scale is large enough - they have hundreds of stores. Previously, one employee could only manage 1-2 stores, but now one person manages 4-5 stores. By saving manpower, it is expected that the cost of model fine-tuning and hardware investment will be recovered next year.

Huang Zhong, the head of an enterprise that has been deeply engaged with all-in-one machines in the past two years, concluded: To use the all-in-one machine, it is often necessary to have a multi-modal configuration of "large model + small model". For example, in the fields of industrial control, traditional manufacturing, etc., the workflow includes multi-modal information input, task planning, deep thinking and command execution, etc., and different models that are good at different tasks need to be equipped. One DeepSeek is obviously not enough - DeepSeek is good at deep thinking rather than command following. It may take five minutes to think about a simple command.

Moreover, Huang Zhong mentioned to Leiphone that the transformation of actual all-in-one machine demand is often slow.

One point is that there is a gap between the model and the application. Decision makers need to fully investigate what application to build with the model before they can see the results. He also observed that many all-in-one machine projects are often for piloting certain scenarios in the early stage. Users do not have high requirements for the number of concurrent users, and private deployment and privacy are the top priority.

On the other hand, there are more fundamental limitations on fine-tuning of large models - even for the open source DeepSeek large model, users cannot intervene in the pre-training stage. For industries such as hospitals with a large amount of local data, it is difficult for data to really enter the model, and the match with the business service is still insufficient. In the eyes of industry insiders, being able to truly load their own data into the large model is an opportunity for products such as all-in-one machines to cultivate user stickiness.

The real realization of the value of all-in-one machines also requires the completion of killer applications and important user data.

Note: In this article, Yang Fei, Jianing, and Huang Zhong are all pseudonyms.