Meta Compute: Everyone Wants To Be A Cloud

元计算：人人皆想成云

Zuck Takes Plan B? SpaceX 2.0, Bedrock 2.0, MSL Isn't Giving Up, Scaling RecSys by 10x... ClusterMAX ranking coming soon?

扎克伯格选择 B 计划？SpaceX 2.0、Bedrock 2.0、MSL 不轻言放弃、推荐系统规模扩大 10 倍……ClusterMAX 排名即将揭晓?

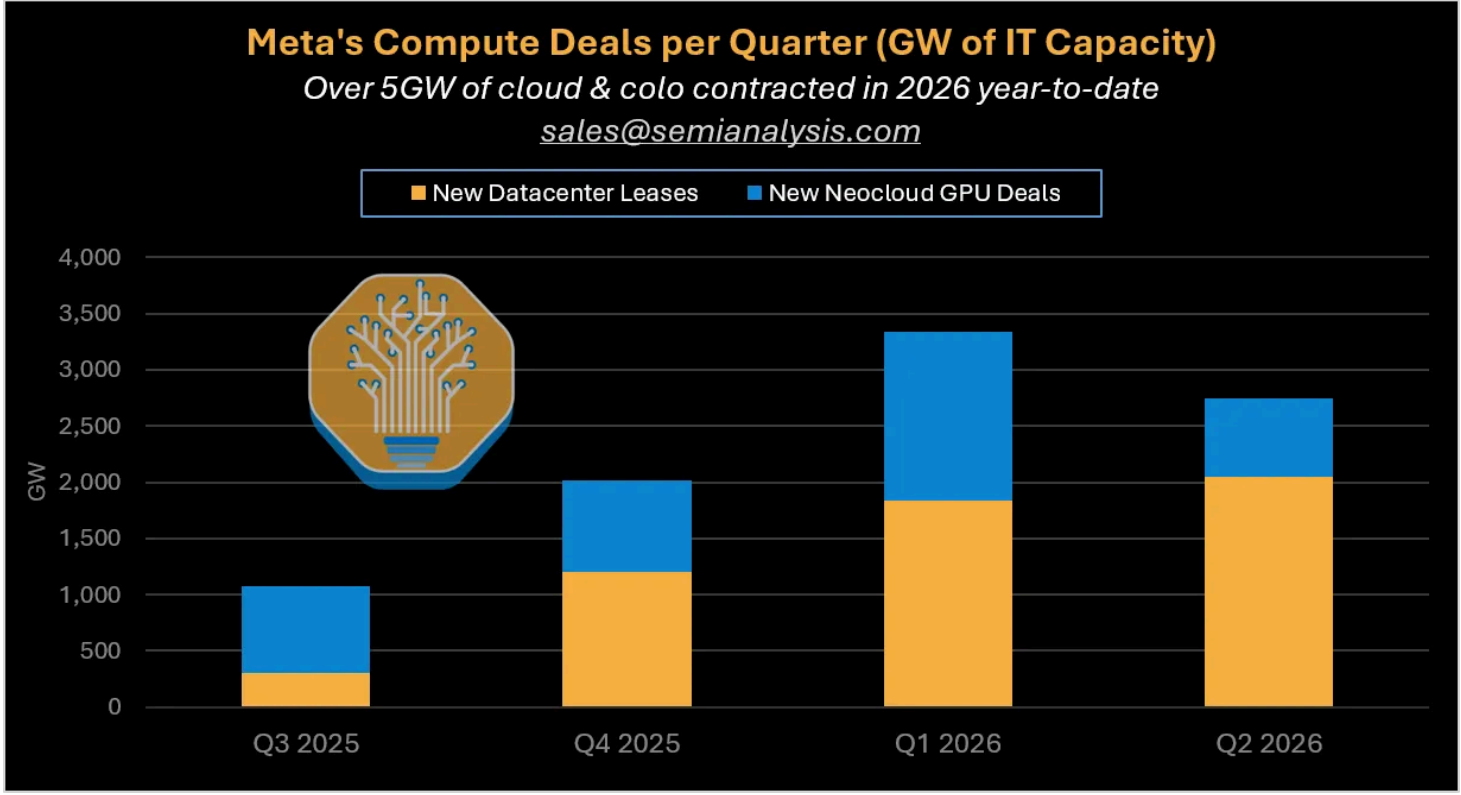
JEREMIE ELIAHOU ONTIVEROS, MAX KAN, JOEY BROOKHART, AND 2 OTHERS

杰雷米·埃利亚乌·昂蒂维洛斯、马克斯·卡恩、乔伊·布鲁克哈特及另外两人

JUL 03, 2026 2026 年 7 月 3 日 · PAID · 已付款

With Bloomberg headlines suggesting Meta could become a Neocloud, the market's reaction was immediate: aggressive sell-off of Neoclouds like Coreweave & Nebius, and debates of "overcapacity" coming back. Let's set the record straight – we believe that **both takes are erroneous** and that Meta's datacenter & compute procurement will ***accelerate***, not slow down. Capex in 2027 will be shockingly high. In just the first six months of the year, **Meta has contracted over 5GW of capacity across Cloud & Colo**, and that doesn't even include all their accelerating self-build activity. Everything is computer and everything is a neocloud.

随着彭博社头条宣称 Meta 可能成为一家 Neocloud，市场反应立竿见影：Coreweave、Nebius 等 Neocloud 公司遭猛烈抛售，"产能过剩"的争论卷土重来。让我们澄清事实——我们认为这两种观点都是错误的，Meta 的数据中心与算力采购只会加速而非放缓。2027 年的资本支出将高得惊人。仅今年上半年，Meta 已在云服务和托管领域签约超过 5GW 的容量，这甚至还未计入其加速推进的自建项目。万物皆算力，万物皆 Neocloud。



Source: SemiAnalysis Datacenter Model

来源：SemiAnalysis 数据中心模型

Meta's capacity under construction just keeps accelerating. We show below Meta's two largest campuses – these two pictures represent 2.5GW of capacity under construction! By the way, if you believed the laughable headlines that "half of US datacenters and delayed and only 5GW are under construction", we show you here that just two datacenters are half of this. Read our piece "Stop Saying Half of US Datacenters Are Delayed" for more on why these headlines are completely off.

Meta 在建的算力规模持续加速增长。下图展示了 Meta 最大的两个园区——这两张图片代表 2.5GW 的在建容量！顺便说一句，如果你相信那些可笑的标题党文章声称"美国半数数据中心延期，仅 5GW 在建"，我们在此展示两个数据中心就占了其中一半。请阅读我们的文章《别再说什么美国半数数据中心延期了》，了解更多关于这些头条为何完全失实的原因。

Source: SemiAnalysis Datacenter Model

来源：SemiAnalysis 数据中心模型

Of course, this naturally raises the questions of what Meta will do with this compute, and whether they're going to flood the market with all of this supply if they turn into a Neocloud. Broadly speaking we see four major high-value use-cases, which are all differentied and very different relative to what traditional Neoclouds do:

当然，这自然会引发一个问题：Meta 将用这些算力做什么？如果他们转型为 Neocloud，是否会向市场大量投放这些算力？总的来说，我们看到了四个主要的高价值用例，它们各有特色，与传统 Neocloud 的做法截然不同：

1. Frontier AI Models: Meta has NOT given up on training frontier models. The bulk of incremental capacity still goes to Meta Superintelligence Labs, and we think the team is currently excited about their progress. A follow-up report will dive into MSL, their unique data strategy, and discuss their chances of catching up with Anthropic and OpenAI. Of course, our Tokenomics Model subscribers

already know our takes and have access to all of this real-time.

前沿 AI 模型：Meta 并未放弃训练前沿模型。绝大部分新增算力仍流向 Meta 超级智能实验室，我们相信该团队目前对自己的进展颇为振奋。后续报告将深入分析 MSL、其独特的数据策略，并探讨他们追赶 Anthropic 和 OpenAI 的可能性。当然，我们的 Tokenomics 模型订阅用户早已了解我们的观点，并可实时获取所有这些信息。

2. RecSys: We believe Meta thinks they can scale up Ads recommendation systems by >10x in complexity to accelerate revenue growth. That requires both inference & training compute for their RecSys models. They can also do more generative targeted ads.

推荐系统：我们认为 Meta 认为他们可以将广告推荐系统的复杂性提升超过 10 倍，以加速收入增长。这需要为其推荐系统模型同时提供推理和训练算力。此外，他们还可以制作更多生成式定向广告。

3. (***SemiAnalysis Exclusive***) We believe that **Meta is in final talks with Anthropic to get access to private instances of Claude**. This would be akin to Bedrock, Foundry, Vertex from other hyperscalers (read our deep dive here). There are multiple use-cases for Meta, ranging from internal usage, to building the premier Sales & Marketing SaaS powered by Frontier AI Agents. We expect Meta to launch a token-as-a-service endpoint and increasingly move up the stack, leveraging its network and distribution. Initially it will be their own models externally and Anthropic internaly, but over time we believe they will serve Anthropic and OpenAI models externally

（SemiAnalysis 独家）我们认为，Meta 正与 Anthropic 进行最终谈判，以获得 Claude 私有化实例的访问权限。这类似于其他超大规模云服务商的 Bedrock、Foundry、Vertex 等产品（详见我们的深度分析）。对 Meta 而言，这存在多重应用场景：从内部使用到打造由前沿 AI 智能体驱动的顶级销售与营销 SaaS 平台。我们预计 Meta 将推出"代币即服务"端点，并借助其网络与分发能力持续向上层渗透。初期可能面向外部提供自有模型、内部部署 Anthropic 模型，但长期来看，我们认为它们将同时对外提供 Anthropic 和 OpenAI 的模型。

4. We expect Meta to strike a few “SpaceX-type” deals. Elon is a sales genius, and he created a **brand new market segment: large-scale on-demand compute at a huge pricing premium**. We think Meta wants in, but selectively. After all,

just a couple hundred MWs can already drive >$10B of yearly revenue! We expect a ten billion dollar Anthropic deal to kick off the flywheel.

我们预计 Meta 将达成数项"SpaceX 式"交易。埃隆堪称销售天才，他开创了一个全新市场领域：以超高定价溢价提供大规模按需算力服务。我们认为 Meta 希望选择性入局——毕竟仅需数百兆瓦的算力就能创造超百亿美元年收入！预计一笔百亿美元级别的 Anthropic 交易将启动这一增长飞轮。

This high optionality, with four high-value-add options, makes it easy for Meta to keep aggressively contracting compute. Meta Superintelligence remains the core engine, but if it doesn't work out, there are many high-margin alternatives to sell compute. It is essentially a CFO's dream and makes it very easy to go all-in on compute – we bet Susan did a 180° flip when she saw the pricing of SpaceX compute deals! Meta won't be a normal bare-metal IaaS vendor with ~30% gross margins – all its options are high value, and enable to easily afford paying a margin to other Neoclouds in order to accelerate their fleet buildout - even if MSL doesn't work out.

这种高可选性（包含四种高附加值选项）让 Meta 能够轻松持续激进采购算力。Meta 超级智能仍是核心引擎，但若此路不通，仍有众多高利润替代方案可供出售算力。这本质上是首席财务官的梦想，让全盘押注算力变得易如反掌——我们打赌苏珊在看到 SpaceX 算力交易的定价后，态度一定发生了 180 度大转弯！Meta 不会成为普通裸金属 IaaS 供应商那样仅有~30%的毛利率——其所有选项都具有高价值，即便 MSL 未能成功，也能轻松承担向其他 Neocloud 支付溢价以加速其集群建设的成本。

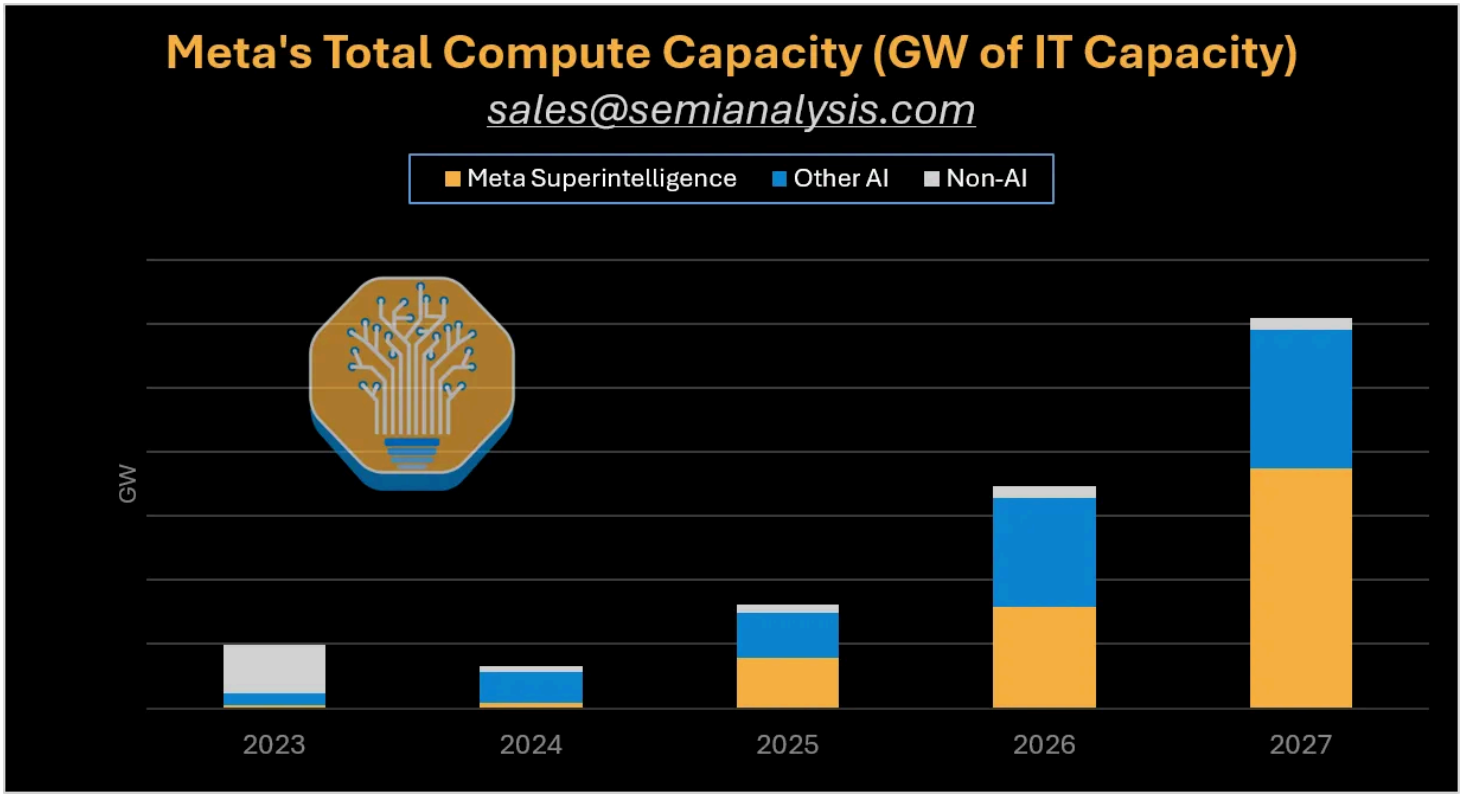
Our Datacenter Model breaks down quarter by quarter their capacity additions across selfbuild, datacenter leasing, and cloud renting. After nearly 10GW of deals signed since early 2024, the bulk of their capacity additions are now through 3rd parties. We expect this to continue and believe that Meta will be a huge source of RPO growth for the likes of Coreweave, Nebius and others.

我们的数据中心模型按季度拆解了 Meta 在自建、数据中心租赁和云租赁方面的产能增长。自 2024 年初以来，Meta 已签署近 10 吉瓦的合同，其大部分新增产能如今通过第三方实现。我们预计这一趋势将持续，并认为 Meta 将成为 Coreweave、Nebius 等公司 RPO（剩余履约义务）增长的重要源泉。

Our model also breaks down, quarter by quarter, Meta's capacity into MSL, other AI, and non-AI. While MSL's huge growth is understandable (they need to keep up with Anthropic & OpenAI!), the surge of Other AI in 2026 suggests plenty of optionality to

monetize compute across RecSys, “SpaceX equivalent”, and Bedrock-type token as a service. If RecSys scales a bit less than expected, they have other options. Of course, if MSL fails, Cloud capacity will skyrocket.

我们的模型同样按季度将 Meta 的算力划分为 MSL（元超级智能实验室）、其他 AI 和非 AI 三大类。尽管 MSL 的迅猛增长合乎情理（其需与 Anthropic 及 OpenAI 保持同步！），但 2026 年其他 AI 领域的激增，预示着 Meta 在推荐系统、SpaceX 级交易、以及 Bedrock 式代币即服务等算力变现方面拥有丰富选择空间。若推荐系统扩展幅度略低于预期，他们仍有备选方案。当然，若 MSL 遭遇挫折，云算力需求将急剧攀升。



Source: SemiAnalysis Tokenomics Model, Datacenter Model

数据来源：SemiAnalysis 代币经济学模型、数据中心模型

Let’s now dig into these four options. We start with the SpaceX-like deal, and the Bedrock-type ambition. Then, behind paywall, we discuss the outlook for MSL and for RecSys.

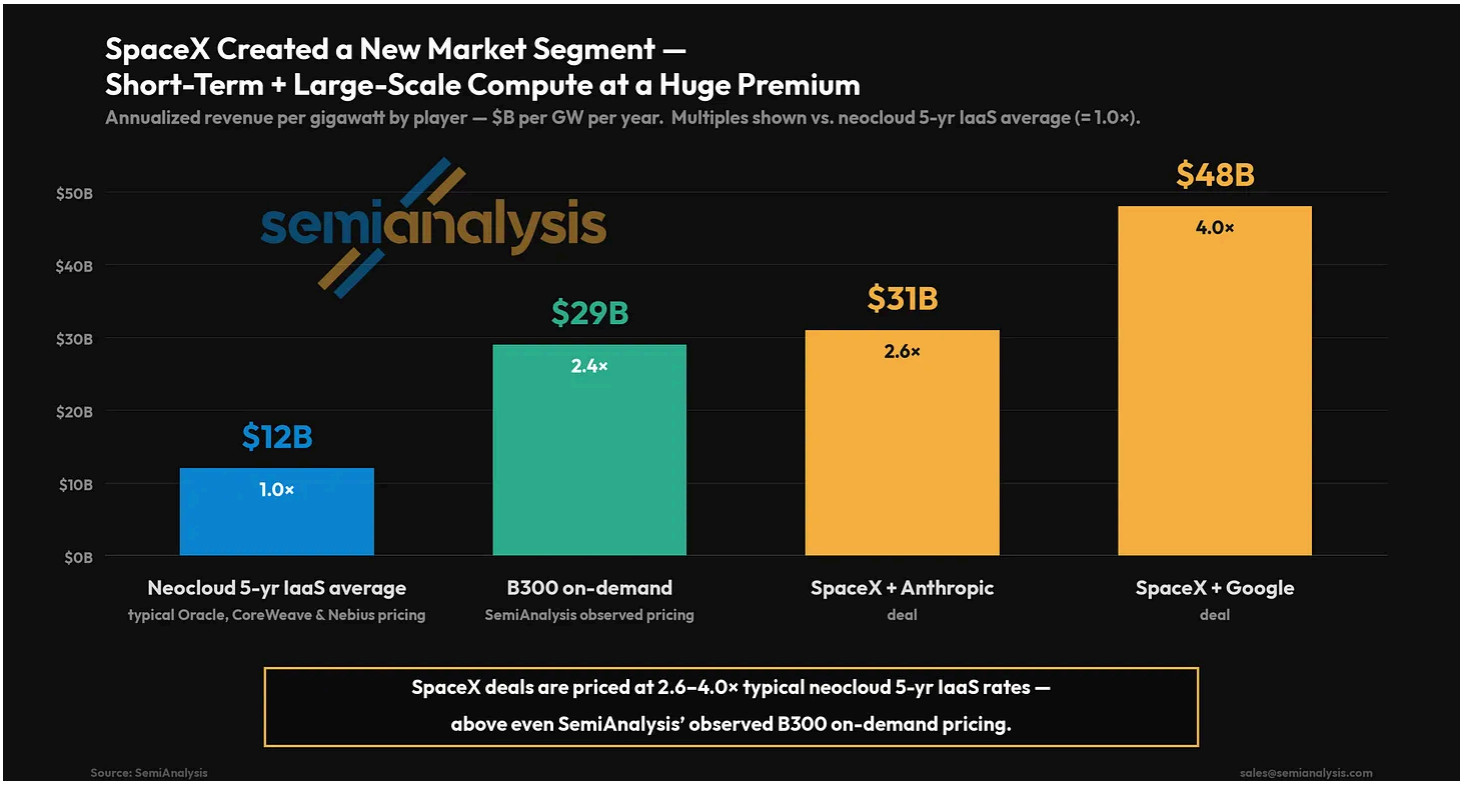
接下来我们将深入探讨这四个选项。首先聚焦于 SpaceX 级交易和 Bedrock 式布局，随后在付费内容部分，我们将分析 MSL 和推荐系统的发展前景。

SpaceX dreams – Elon's genius deals

SpaceX 梦想——埃隆的天才交易

Elon Musk shocked the AI infrastructure world when he announced the first Anthropic deal. He double-shocked the world when he came up with the Google deal. The reason is simple: the revenue per MW of these deals are respectively triple and quadruple what peers are charging, as shown below. The profit per MW gap is, of course, even larger given the cost structure is roughly the same.

埃隆·马斯克宣布与 Anthropic 的首笔交易时震惊了 AI 基础设施界，而当他达成谷歌的交易时更是让全球为之双重震惊。原因很简单：如下图所示，这些交易的每兆瓦营收分别是同行收费的三倍和四倍。考虑到成本结构大致相同，每兆瓦利润的差距自然更为悬殊。



Source: AI Cloud TCO Model

源：AI 云总拥有成本模型

SpaceX's pricing on the Google deal is even higher than what we observe on the on-demand or short-term rental market. Elon essentially invented a new market segment. Our AI Cloud TCO model team tracks hundreds of GPU cloud deals every year, including all contract terms like SLA, pricing, contract length etc. That's why we have,

by far, the world's best information on GPU pricing by contract length – see a preview in our free dashboard here.

SpaceX 在谷歌交易中的定价甚至高于我们观察到的按需或短期租赁市场价格。埃隆本质上创造了一个全新的细分市场。我们的人工智能云 TCO 模型团队每年追踪数百笔 GPU 云交易，涵盖 SLA、定价、合同期限等所有合同条款。正因如此，我们拥有目前全球最完善的按合同期限划分的 GPU 定价信息——您可通过我们的免费仪表盘预览相关内容。

We have never ever seen a deal that large, and that short: the deal is three years, but with option to cancel from both parties within 90 days – so it's effectively a 3-month deal with automatic renewal.

我们从未见过规模如此庞大且期限如此短暂的交易：合同期为三年，但双方均可在 90 天内取消——因此这本质上是一份三个月自动续约的协议。

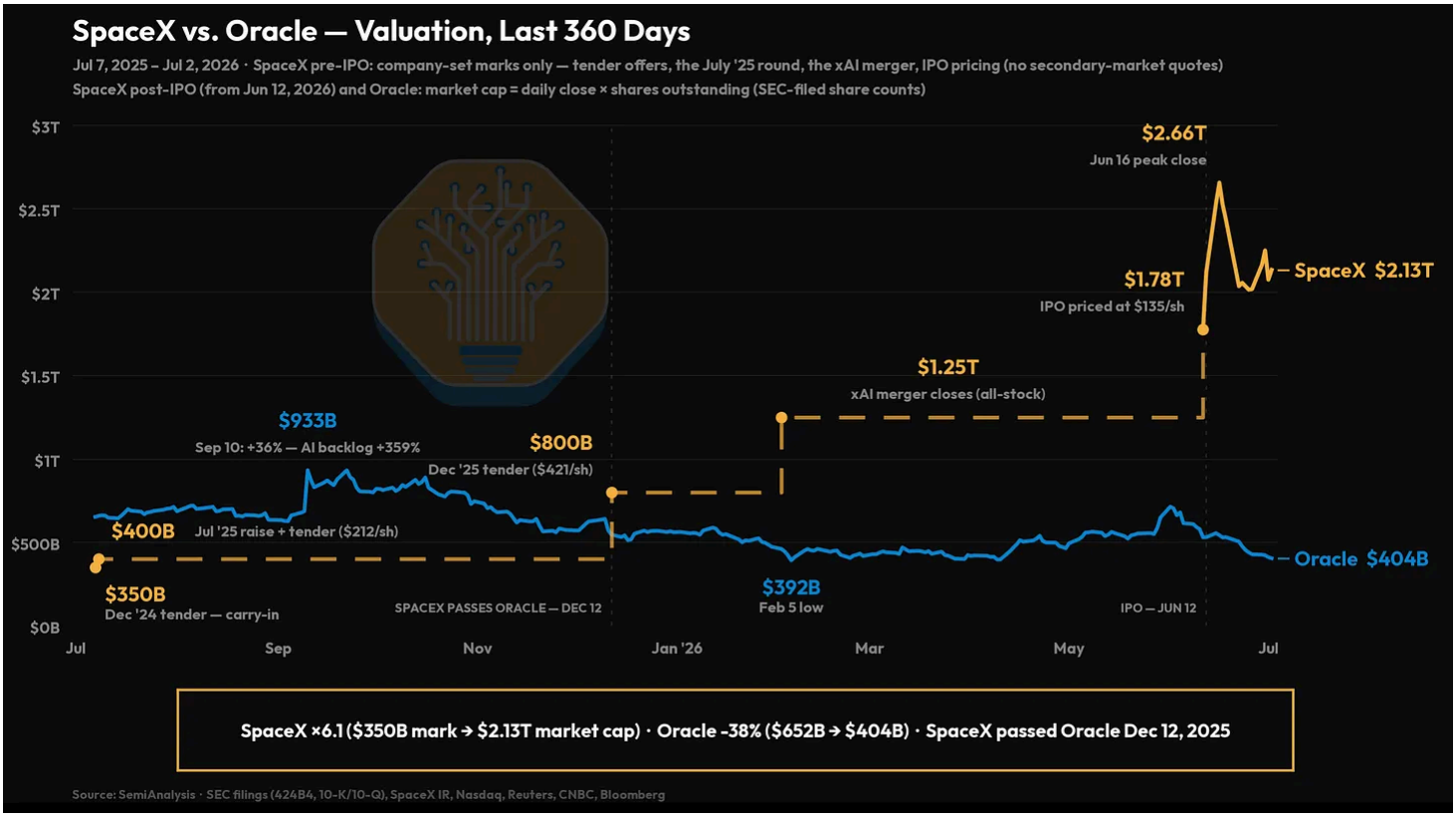
The reason this has never happened is because very few companies can do it. The financing burden excludes all the Neoclouds from participating in that market – for large clusters, they need to secure an offtaker over multiple years. The top 3 hyperscalers could've done it… but they all see higher-value long-term options, such as Microsoft getting OpenAI's IP in exchange for equity investments & compute, Amazon focused on increasing the adoption of Bedrock and Trainium, Google similarly with TPUs and Vertex (now Gemini Enterprise Platform).

之所以从未出现过这类交易，是因为鲜有企业能达成。融资压力将所有 Neocloud 企业排除在该市场之外——对于大型集群而言，它们需要锁定多年期的承购方。三大超大规模云服务商本可做到……但它们都看到了更高价值的长期选项，例如微软通过股权投资及算力获取 OpenAI 知识产权，亚马逊专注提升 Bedrock 和 Trainium 采用率，谷歌则在 TPU 和 Vertex（现为 Vertex Enterprise AI 平台）领域如法炮制。

This effectively leaves only two companies that can truly capitalize on this new market: Oracle and Meta. For the former, we see this as a massive blow. It's yet another piece of evidence that they could've done a much better job at monetizing all of their gigawatts of compute. A look at the valuation trajectory of Oracle vs SpaceX illustrates the massive divergence. For both, Gigawatts are an increasing portion of

their valuation, and a large driver of that evolution within the last year.

这实际上只剩下两家公司能真正从这个新市场中获利：甲骨文与 Meta。对前者而言，这无疑是一记重创——再次证明他们本可以更好地将数十亿瓦计算资源变现。对比甲骨文与 SpaceX 的估值轨迹，便能看出巨大分野：对两者而言，千兆瓦级算力在估值中的占比都在攀升，且成为近一年来估值变化的主要推手。



Source: SemiAnalysis, SEC, CNBC/Bloomberg/Reuters

来源：SemiAnalysis、美国证券交易委员会、CNBC/彭博社/路透社

Enter Meta. Given their laser-focus on building Superintelligence, they probably didn't put enough thought into the Cloud option. SpaceX and Elon paved the way for them. At $50B/Gigawatt of annual revenue, it becomes an easy call. Just allocating 200MW of compute to an external customer drives $10B/yr of revenue, at sky-high margin. And the ability to cancel that contract within 90 days makes it even easier – if they want to give MSL more compute, they can, in short notice.

再看 Meta。由于他们全力聚焦超级智能研发，可能对"云端选项"缺乏足够思考。SpaceX 与马斯克为他们开辟了道路。按每年每千兆瓦 500 亿美元营收计算，决策变得轻而易举——仅需分配 200 兆瓦计算资源给外部客户，就能带来每年 100 亿美元营收，且利润率极高。而可在 90 天内取消合约的灵活性更让决策毫无包袱：若 Meta 计划为超级智能实验室(MSL)调配更多算力，也能迅速调整。

That's also a perfect match for Meta's datacenter construction strategy. A year ago, we were the first to track their new "tent" ultra-fast datacenter design. Since then, Meta tents have been popping up all over the US! By bringing datacenters online fast, even if "lower quality", it'll be easier for Meta to monetize compute in a SpaceX fashion.

这也与 Meta 的数据中心建设策略完美契合。一年前，我们率先追踪到他们新型"帐篷式"超快速数据中心设计。此后，Meta 的帐篷式数据中心如雨后春笋般遍布全美！通过快速上线数据中心——即便"质量稍逊"——Meta 将更容易以 SpaceX 式的方式实现算力变现。

Source: SemiAnalysis Datacenter Model

来源：SemiAnalysis 数据中心模型

As such, we expect Meta to announce such a deal soon. It's a great way to get the flywheel started. Anthropic is our prime suspect, but others could step in as well, such as OpenAI or Google.

因此，我们预计 Meta 很快会宣布此类合作。这是启动飞轮效应的绝佳方式。Anthropic 是我们的首要怀疑对象，但 OpenAI 或谷歌等其他公司也可能介入。

Will Meta build the new Bedrock?

Meta 会打造新一代 Bedrock 吗?

Another great option for Meta is to strike a deep partnership with frontier labs to sell their models with Meta's own compute. As said earlier, we believe they are in final talks with Anthropic to get private access to their LLMs, just like Amazon has through Bedrock (agreement explained in detail here). This means that for Meta, another way to monetize their compute will be to sell Claude. We see three main paths:

对 Meta 而言，另一个绝佳选择是与前沿实验室建立深度合作，通过 Meta 自身的算力销售其模型。如前所述，我们相信 Meta 正与 Anthropic 进行最终谈判，以获取其 LLMs 的私有访问权限——正如亚马逊通过 Bedrock 所实现的那样（详见协议解读）。这意味着对 Meta 来说，另一种变现算力的方式将是销售 Claude。我们看到了三条主要路径：

1. Part of it could be for internal usage only. Meta needs Claude tokens, and Anthropic can't keep up with demand. There are also security & privacy layers. In the future, we could see other very large enterprises striking such deals, e.g. a JPMorgan probably won't go all-in on Claude if they don't have the security and privacy guarantees that private instances enables (in their own datacenters).

   部分可能仅用于内部用途。Meta 需要 Claude 的令牌（token），而 Anthropic 无法满足需求。此外还涉及安全与隐私层级。未来，我们可能看到其他超大型企业也达成此类合作，例如摩根大通若无法获得私有实例（在其自有数据中心内）所提供的安全与隐私保障，就不太可能全面押注 Claude。

2. Meta could sell Claude as a service in the same way Bedrock does. They have capacity and own the full stack, from CPU to GPU to networking, with high security. However, as a new entrant, it'll be tough to create all the enterprise relationships that AWS has. But Meta has the option to leverage their customer base of advertisers and better integrate frontier agents and LLMs into their suite,

creating a new large-scale distribution path.

> Meta 可以像 Bedrock 那样，将 Claude 作为一项服务进行销售。他们拥有充足的算力，并掌控着从 CPU 到 GPU 再到网络的完整技术栈，安全性极高。然而，作为新入局者，要像 AWS 那样建立所有企业级合作关系将非常困难。但 Meta 可以选择利用其广告客户基础，更好地将前沿智能体和 LLMs 集成到其产品套件中，从而开辟一条全新的大规模分发路径。

3. Meta could go even more vertical and start building applications. As one of the world's largest advertising platforms, they have a path to building a Sales & Marketing powerhouse. Integrating frontier models and agents would increase their odds of building a world class solution.

> Meta 甚至可以更垂直地发展，开始构建应用。作为全球最大的广告平台之一，他们有机会打造一个销售与营销的超级引擎。整合前沿模型和智能体将增加他们构建世界级解决方案的胜算。

There is, of course, also the opportunity to distribute models to free social media users, and to the broader Meta ecosystem including things like connected glasses. It appears more likely that they will prioritize use of their own models, but it's good to keep the optionality, and they likely have a great path to monetization. The huge distribution potential and network effect means that the likes of OpenAI and Anthropic are likely seeing this as highly strategic, and ready to make concessions to get a piece of the pie.

> 当然，还有机会将模型分发给免费社交媒体用户，以及更广泛的 Meta 生态系统（包括智能眼镜等产品）。他们似乎更倾向于优先使用自家模型，但保留选择权也是明智之举，而且很可能有很好的变现路径。巨大的分发潜力和网络效应意味着，像 OpenAI 和 Anthropic 这样的公司很可能将其视为高度战略性的机会，并愿意做出让步以分得一杯羹。

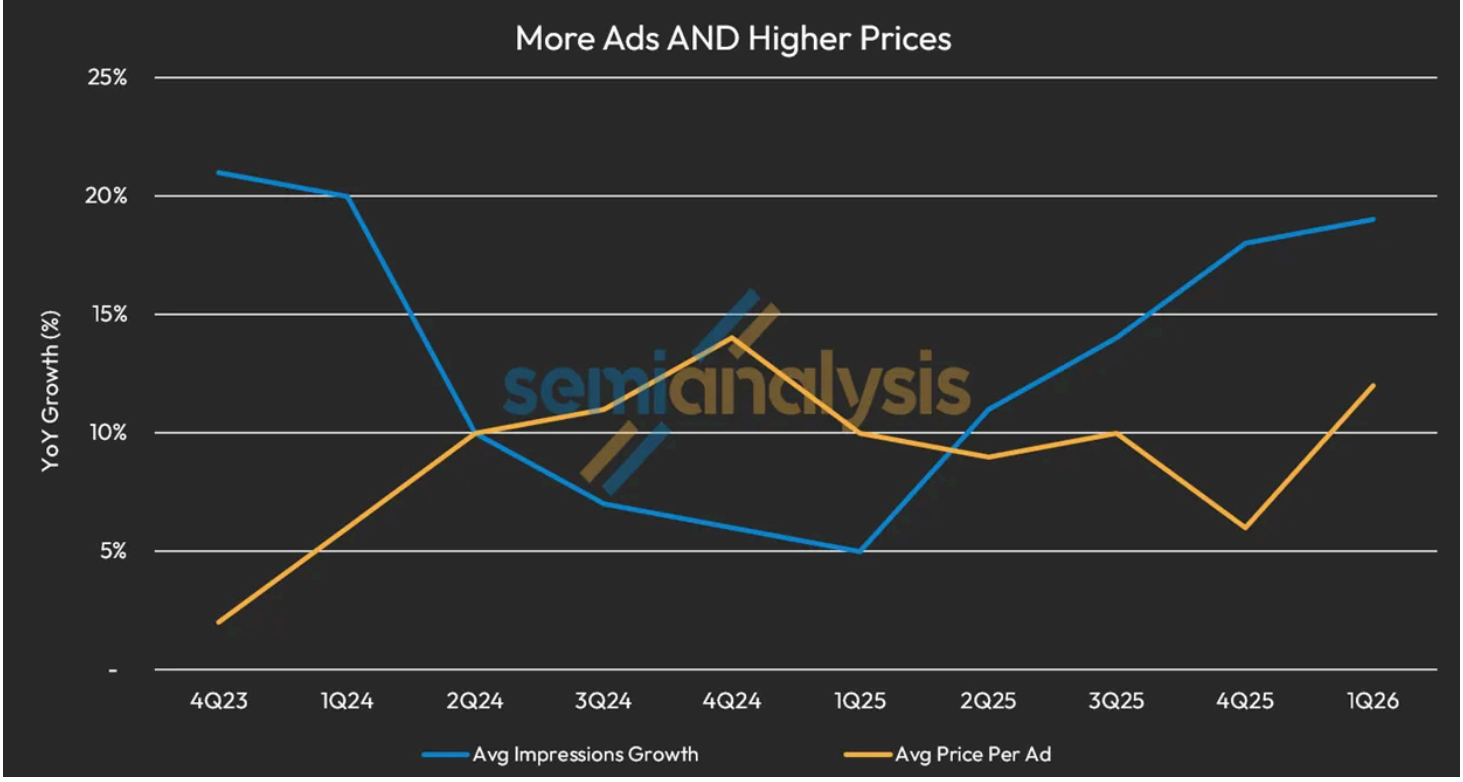
Scaling Ads Recommendation systems: how much can Meta

accelerate?

扩展广告推荐系统：Meta 能加速到什么程度?

For Meta, the key AI story is how RecSys is contributing to revenue acceleration at massive scale. At the end of 2022 and early 2023, the market widely saw the company as entering a stage of maturity and low growth amongst an investment cycle. Today, it's crystal clear that Meta has dramatically re-accelerated revenue growth, in large part due to GPU investments. They have been a key trigger, both on training and inference: RecSys models for ads have been getting bigger and more expensive to run, but much smarter, driving better yields for advertisers as seen in the ability for advertisers to pay higher prices while still seeing strong ROAS. At the same time the RecSys models for content have led to more time on platform across the Family of Apps, improving the monetizable surface area and leading to strong ad imprssions growth.

对于 Meta 而言,AI 领域的核心故事在于推荐系统如何大规模推动收入加速增长。在 2022 年底至 2023 年初,市场普遍认为该公司在投资周期中已进入成熟期和低增长阶段。然而如今已清晰可见,Meta 戏剧性地重新加速了收入增长,这在很大程度上归功于 GPU 投资。无论是训练还是推理环节,GPU 都成为关键驱动因素:用于广告的推荐系统模型规模不断扩大、运行成本持续攀升,但智能化水平显著提升——这从广告主愿意支付更高单价的同时仍能获得强劲的广告支出回报率可见一斑。与此同时,内容推荐系统模型促使全家桶应用的用户使用时长增长,不仅扩大了可变现的流量池,更带来广告展示量的强劲增长。



Source: SemiAnalysis Tokenomics Model

来源: SemiAnalysis 代币经济学模型

The key question on everyone's mind is now – is this sustainable? How much can Meta accelerate core revenue growth? While Meta is playing catching up on the frontier lab side, the non-Meta SuperIntelligence AI chip fleet is producing outstanding ROI. We cover this split compute split between RecSys AI, Meta SuperIntelligence, and other Meta LLM products in our Tokenomics Model.

大家最关心的问题是——这种模式能否持续？Meta 能将其核心收入增长加速到什么程度？虽然 Meta 在前沿实验室领域仍在追赶，但其非 Meta 超级智能 AI 芯片集群已经产生了卓越的回报。我们在 Tokenomics 模型中涵盖了 RecSys AI、Meta 超级智能和其他 Meta LLM 产品之间的这种算力分配。

Behind paywall, we discuss the outlook for RecSys scaling, and then for Meta Superintelligence Labs.

付费墙后，我们将探讨推荐系统扩展的前景，以及 Meta 超级智能实验室的未来。

Prior to GEM, ad ranking ran on DLRMs which don't follow the same exponential power laws seen in LLM and therefore failed to scale with compute. Meta's generative recommender work, Hierarchical Sequential Transduction Units (HSTU), changes that by reformulating ranking as a sequential prediction problem that scales in performance with compute. It beat prior baselines by ~66% on ranking metrics and is

productionized as GEM, Meta's ads foundation model. Benefits of GEM includes:

在 GEM（通用推荐模型）之前，广告排序依赖 DLRM（深度学习推荐模型），这类模型不遵循 LLM（大语言模型）中常见的指数幂定律，因此无法随算力增长而扩展。Meta 的生成式推荐器工作——分层序列转导单元（HSTU）——通过将排序重构为随算力提升而扩展性能的序列预测问题，改变了这一局面。它在排序指标上比之前基准提升了~66%，并已产品化为 Meta 的广告基础模型 GEM。GEM 的优势包括：

· A scalable model that is 4x better at driving ad performance gains and compute efficient per unit of performance gain

一个可扩展的模型，在推动广告效果提升方面表现优异 4 倍，且每单位性能增益的计算效率更高

· A new training stack that has 23x effective training FLOPs with ~1.4x increase in MFU using 16x more GPUs

· 一套新的训练堆栈，其有效训练 FLOPs 提升 23 倍，在 MFU 提高 1.4 倍的同时，GPU 数量增加 16 倍。

We think Meta can profitably absorb a >10x increase in AdRec compute with this new model because each additional gigawatt correlates directly to better predictions, both on the training and inference side. This improvement is showing up through an increase in impressions delivered and average price per ad. In Q1 2026, ad impressions grew 19% YoY and average price per ad grew 12% YoY. This is a continuation CFO Susan Li has attributed their increased average price per ad to improved ad performance and targeting capabilities. After doubling GEM's training GPUs, ad conversion rate on Instagram increased on both Instagram and Facebook by 5% and 3%, respectively.

我们认为，随着这套新模型的引入，Meta 能够盈利地吸收超过 10 倍的广告推荐计算量增长。因为每增加一千兆瓦的计算能力，都会直接转化为更精准的预测——无论是训练端还是推理端。这种改善体现在广告展示量和每条广告平均价格的提升上。2026 年第一季度，广告展示次数同比增长 19%，每条广告平均价格同比增长 12%。这一趋势延续了首席财务官苏珊·李此前归因于广告效果与定向能力提升所带来的价格增长。在将 GEM 训练 GPU 数量翻倍后，Instagram 和 Facebook 上的广告转化率分别提升了 5%和 3%。

What makes the growth sustainable is that the higher prices are justified because ROAS is growing directly with CPMs. Meta Advantage+ Shopping delivers 32% higher ROAS with 17% lower cost/action relative to manual campaigns. Meta can grow ad revenue in one of two ways: (1) by showing more ads, or (2) charge more per ad. But there's a limit to how many ads it can show in a users' feed before it does more harm to the user's experience than good to ad revenue growth. As long as ROAS continues to grow with price, which has shown to be the case, Meta can keep lifting yield even as ad impressions growth eventually start coming up against ad-load ceiling.

持续增长的关键在于，广告价格上涨的合理性源于 ROAS（广告支出回报率）与 CPM（千次展示成本）同步提升。Meta Advantage+购物广告相较手动广告活动，可实现 32%更高的 ROAS 和 17%更低的单次转化成本。Meta 提升广告收入的途径有两种：一是增加广告展示量，二是提高单次广告价格。但在用户信息流中过度展示广告，会损害用户体验，反而阻碍收入增长。只要 ROAS 能随价格上涨而持续增长（事实证明确实如此），即便广告展示量逼近广告负载上限，Meta 仍能持续提升广告收益。

Meta Superintelligence labs – another dropout, or a real threat to Anthropic and OpenAI ?

Meta 超级智能实验室——又一个中途退场者，抑或是对 Anthropic 和 OpenAI 的真正威胁?

Compute is the lifeblood of every AI research org. Anthropic would never sell compute to any of its competitors, and it's therefore tempting to view Meta's decision as admitting defeat for MSL.

算力是每个 AI 研究机构的命脉。Anthropic 绝不会向任何竞争对手出售算力，因此 Meta 此举很容易被视为 MSL 战略失败的信号。

However, we believe this framing is too simplistic. There is a world where you can temporarily sell compute to your fiercest competitors AND still become a true frontier lab. Like so much else we've covered in this article, Zuck really needs to thank Elon for

showing him this playbook.

然而，我们认为这种框架过于简单化。存在一种可能：你可以暂时将算力出售给最激烈的竞争对手，同时仍能成为真正的前沿实验室。正如本文中讨论的其他许多内容一样，扎克伯格真的需要感谢埃隆向他展示了这套打法。

We believe the following is the most likely explanation for the SpaceX/Cursor acquisition:

我们认为以下是对 SpaceX/Cursor 收购案最合理的解释：

- Elon and Cursor are both still serious about building superintelligence. Cursor's messaging over the past 6 months has changed from "we want to build the best AI coding experience" to "we want to build RSI". This is the ethos of a frontier lab, not a wrapper startup.

  埃隆和 Cursor 双方都仍认真致力于构建超级智能。过去六个月里，Cursor 的叙事已从"我们想打造最佳 AI 编程体验"转变为"我们想构建 RSI"。这是前沿实验室的核心理念，而非包装型初创公司的逻辑。

- SpaceX still has up to 900MW left after their eye-watering compute deals. This is sufficient for taking a few true shots on goal at reaching the frontier. Both GPT 5.6 and Fable 5 were likely trained using less compute.

  SpaceX 在完成其令人瞠目的算力交易后，仍剩余高达 900 兆瓦的电力储备。这足以支撑他们真正向前沿领域发起几次冲击。GPT 5.6 和 Fable 5 的训练很可能使用的算力都未超过这个规模。

- The 90-day bilateral cancellation in all the SpaceX compute deals is crucial! Many were viewing this as a win for Anthropic/Google---because they can just cancel the deal if they find something cheaper---but it also means Elon can claw back all his compute if the Cursor team does well!

  所有 SpaceX 算力交易中的 90 天双向取消条款至关重要！许多人曾认为这对 Anthropic/Google 有利——因为他们若找到更便宜的方案就能直接取消合约——但这同样意味着，如果 Cursor 团队表现优异，Elon 可以收回全部算力资源！

- In other words, SpaceX/Cursor still meet the minimum requirements for training a frontier model. Over the next 6 months, we will see if they can deliver on that

potential. If they do, they will claw back all their compute and continue scaling to push the frontier. If they fail, then they'll give up for real and become a very large and profitable CSP instead. That's certainly not a bad backup plan, and the only true cost will have been the unmonetized GPU hours spent chasing the dream of RSI.

换言之，SpaceX/Cursor 仍满足训练前沿模型的最低门槛。未来六个月将检验他们能否兑现这份潜力。若能成功，他们便会收回所有算力并持续扩大规模以突破边界；若失败，则将彻底放弃转型，成为规模庞大且盈利丰厚的云服务提供商。这当然不算糟糕的备用方案，唯一的真实成本不过是那些追逐 RSI 梦想时未能变现的 GPU 运行时数。

A future post will cover all of this in more detail, but we think Meta is taking the same approach with MSL. The only difference is that they have even more options to monetize the excess compute in the meantime! RecSys and TaaS compute can essentially be repurposed back to research at any time.

后续文章将更详尽地阐述这点，但我们认为 Meta 对 MSL 采取了相同策略。唯一的区别在于，他们拥有更多将冗余算力变现的途径！推荐系统与 TaaS 算力本质上可随时重新调配用于研究工作。

Compute, data, and talent are the three pillars of a frontier lab, and unlike the other hyperscalers, Zuck has put up the capital to try becoming world-class at all three. Success is far from guaranteed, but it'd be a mistake to count them out. The one thing that'd change our minds is if they sign a compute deal without any sort of SpaceX-style early cancellation clause. If that happens, then MSL is actually cooked.

算力、数据与人才是前沿实验室的三大支柱，与其他超级云巨头不同，扎克伯格已投入重金力求在这三方面都达到世界级水平。虽然成功远非板上钉钉，但轻视他们无疑是错误的。唯一可能改变我们看法的，是他们签署一项不含 SpaceX 式提前终止条款的算力交易——若真如此，MSL 便已回天乏术。

Recommend SemiAnalysis to your readers

向您的读者推荐 SemiAnalysis

Bridging the gap between the world's most important industry, semiconductors, and business.

连接全球最重要的半导体产业与商业之间的桥梁。

Recommend 推荐

69 Likes 69 个赞 · 1 Restack 1 次整理

Previous 上一篇

Discussion about this post

关于此帖的讨论

Comments 评论 Restacks

Write a comment...

Eddie 埃迪 Eddie 埃迪 1h

100% agree. Thanks for such a quick and timely piece.

完全同意。感谢您如此迅速且及时地发表这篇文章。

LIKE 点赞 REPLY 回复 SHARE 分享