

O

O

# Trends – Artificial Intelligence

BOND  
May 2025

# Trends – Artificial Intelligence

BOND  
May 2025

Trends – Artificial Intelligence (AI)

May 30, 2025

Mary Meeker / Jay Simons / Daegwon Chae / Alexander Krey

BOND

Trends – Artificial Intelligence  
(AI)

2025 年 5 月 30 日

Mary Meeker / Jay Simons / Daegwon Chae / Alexander Krey

BOND

We set out to compile foundational trends related to AI. A starting collection of several disparate datapoints turned into this beast. As soon as we updated one chart, we often had to update another – a data game of whack-a-mole... a pattern that shows no sign of stopping...and will grow more complex as competition among tech incumbents, emerging attackers and sovereigns accelerates.

Vint Cerf, one of the ‘Founders of the Internet,’ said in 1999, ‘...they say a year in the Internet business is like a dog year – equivalent to seven years in a regular person's life.’ At the time, the pace of change catalyzed by the internet was unprecedented.

Consider now that AI user and usage trending is ramping materially faster...and the machines can outpace us.

The pace and scope of change related to the artificial intelligence technology evolution is indeed unprecedented, as supported by the data. This document is filled with user, usage and revenue charts that go up-and-to-the-right... often supported by spending charts that also go up-and-to-the right.

Creators / bettors / consumers are taking advantage of global internet rails that are accessible to 5.5B citizens via connected devices; ever-growing digital datasets that have been in the making for over three decades; breakthrough large language models (LLMs) that – in effect – found freedom with the November 2022 launch of OpenAI’s ChatGPT with its extremely easy-to-use / speedy user interface.

In addition, relatively new AI company founders have been especially aggressive about innovation / product releases / investments / acquisitions / cash burn and capital raises. At the same time, more traditional tech companies (often with founder involvement) have increasingly directed more of their hefty free cash flows toward AI in efforts to drive growth and fend off attackers.

And global competition – especially related to China and USA tech developments – is acute.

The outline for our document is on the next page, followed by eleven charts that help illustrate observations that follow.

We hope this compilation adds to the discussion of the breadth of change at play – technical / financial / social / physical / geopolitical. No doubt, people (and machines) will improve on the points as we all aim to adapt to this evolving journey as knowledge – and its distribution – get leveled up rapidly in new ways.

Special thanks to Grant Watson and Keeyan Sanjasaz and BOND colleagues who helped steer ideas and bring this report to life. And, to the many friends and technology builders who helped, directly or via your work, and are driving technology forward.

We set out to compile foundational trends related to AI. A starting collection of several disparate datapoints turned into this beast. As soon as we updated one chart, we often had to update another – a data game of whack-a-mole... a pattern that shows no sign of stopping...and will grow more complex as competition among tech incumbents, emerging attackers and sovereigns accelerates.

Vint Cerf, one of the ‘Founders of the Internet,’ said in 1999, ‘...they say a year in the Internet business is like a dog year – equivalent to seven years in a regular person's life.’ At the time, the pace of change catalyzed by the internet was unprecedented.

Consider now that AI user and usage trending is ramping materially faster...and the machines can outpace us.

与人工智能技术发展相关的变化步伐和范围确实是前所未有的，数据也支持了这一点。本文档中充满了用户、使用<sup>a</sup>和收入图表，这些图表呈持续上升趋势 right...，通常还有支出图表也呈上升趋势。

创造者 / 赌徒 / 消费者正在利用全球互联网轨道，55 亿公民可以通过连接设备访问这些轨道；超过三十年历史的不断增长的数字数据集；突破性的大型语言模型 (LLM) – 实际上 – 随着 2022 年 11 月 OpenAI 的 ChatGPT 及其极其简单的易于使用 / 快速的用户界面的推出而获得了自由。

In addition, relatively new AI company founders have been especially aggressive about innovation / product releases / investments / acquisitions / cash burn and capital raises. At the same time, more traditional tech companies (often with founder involvement) have increasingly directed more of their hefty free cash flows toward AI in efforts to drive growth and fend off attackers.

And global competition – especially related to China and USA tech developments – is acute.

我们的文档大纲在下一页，后面是 11 张图表，用于说明以下观察结果。

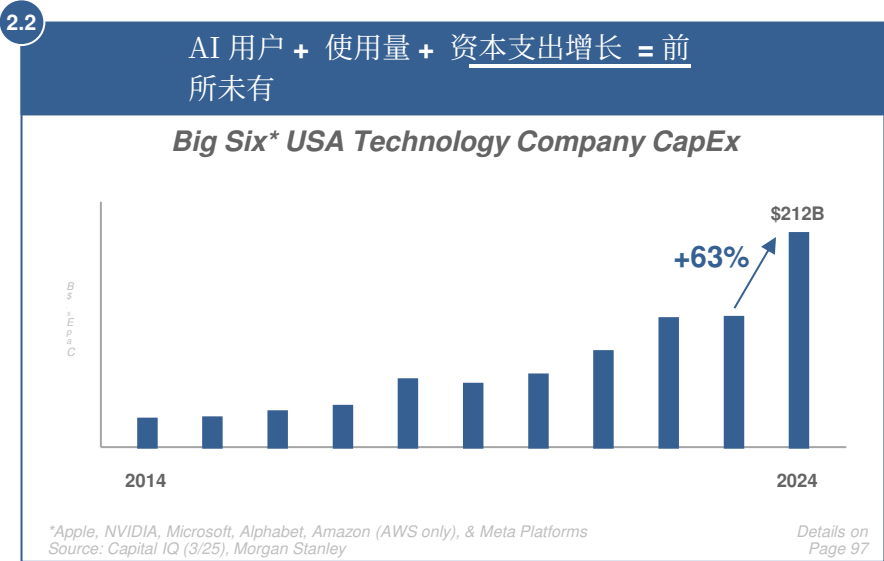
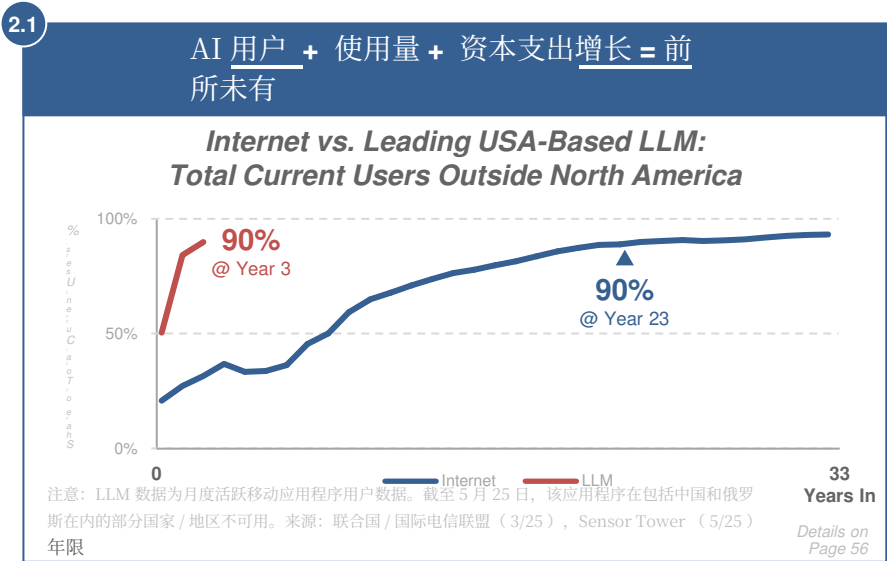
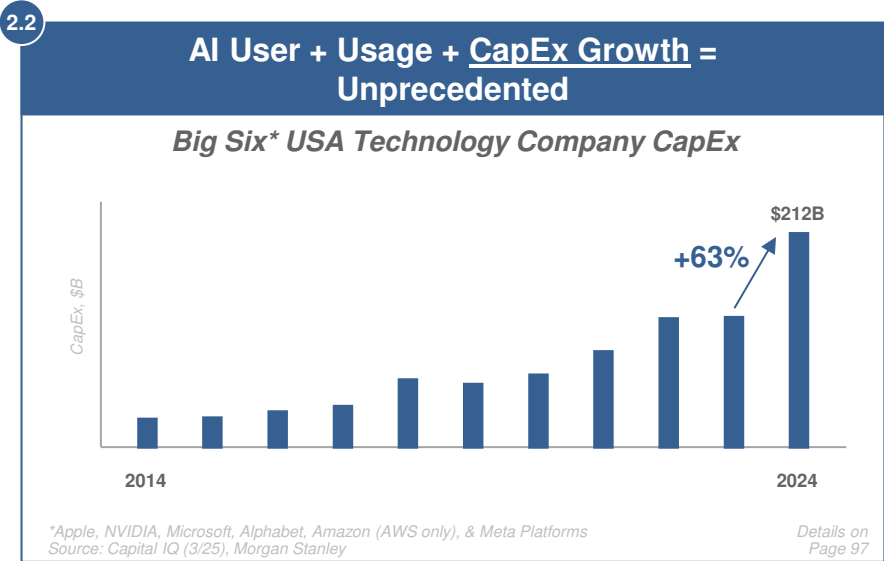
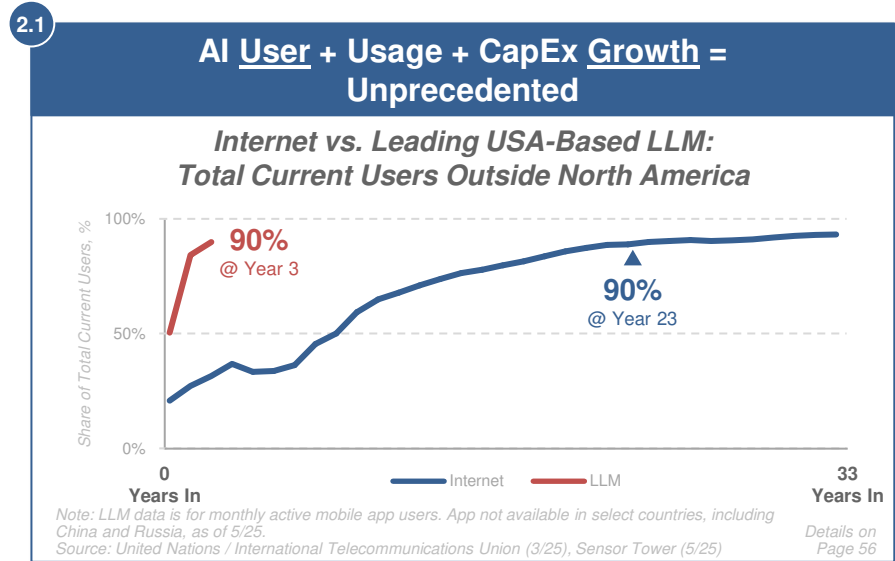
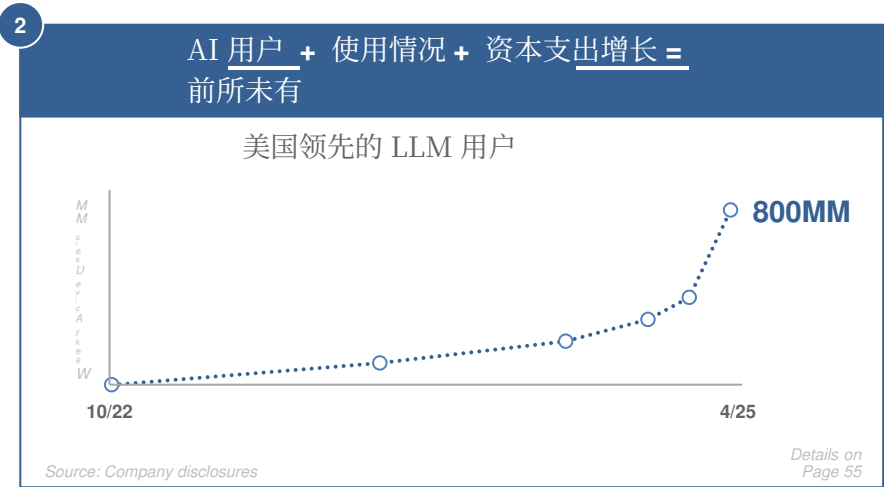
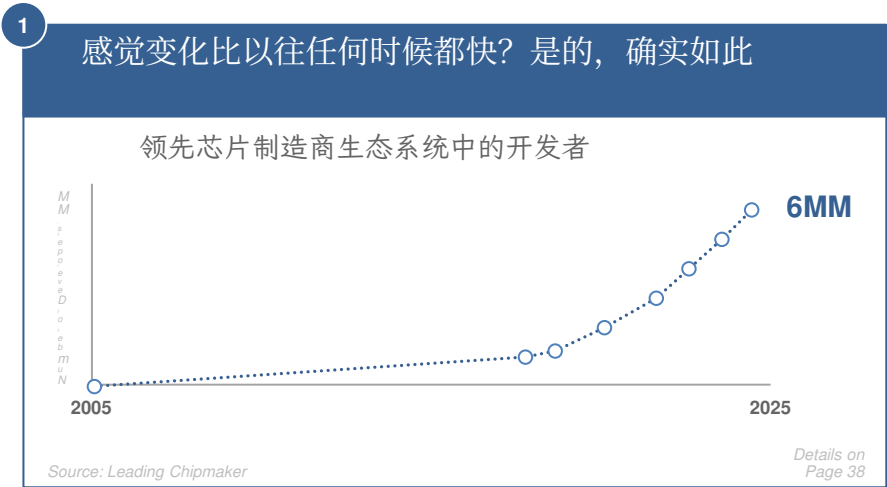
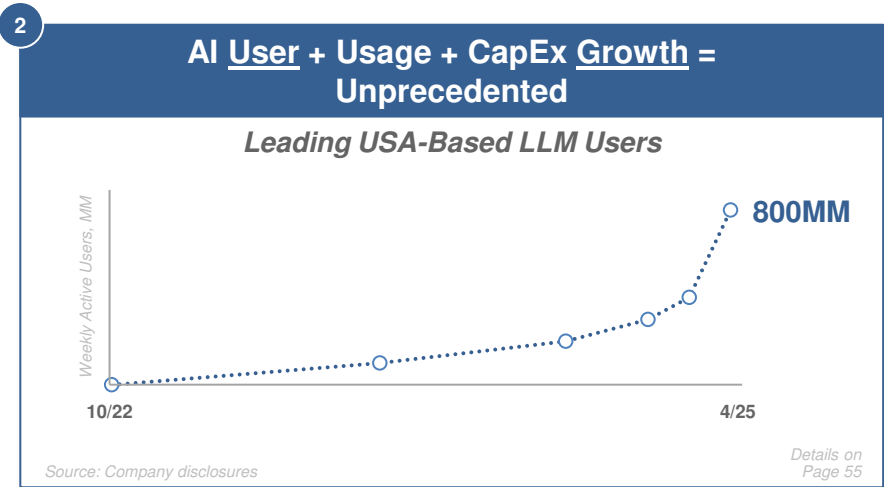
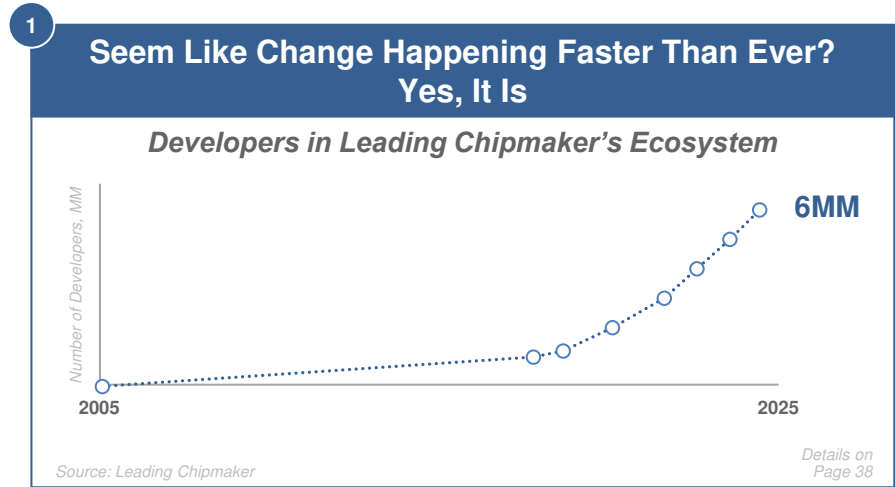
我们希望这份汇编能够促进对正在发挥作用的变革广度的讨论 – 技术 / 金融 / 社会 / 物理 / 地缘政治。毫无疑问，人们（和机器）将改进这些要点，因为我们都旨在适应这种不断发展的旅程，因为知识 – 及其传播 – 以新的方式迅速提升。

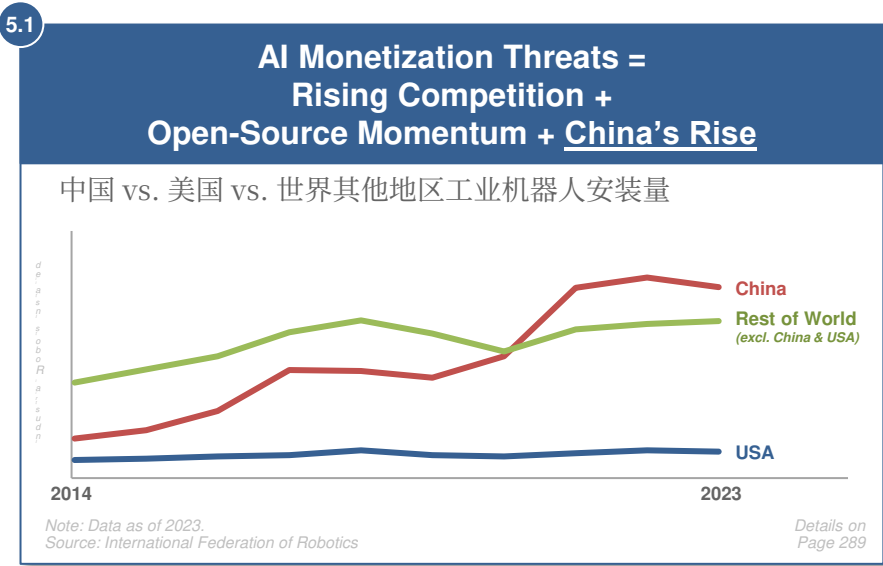
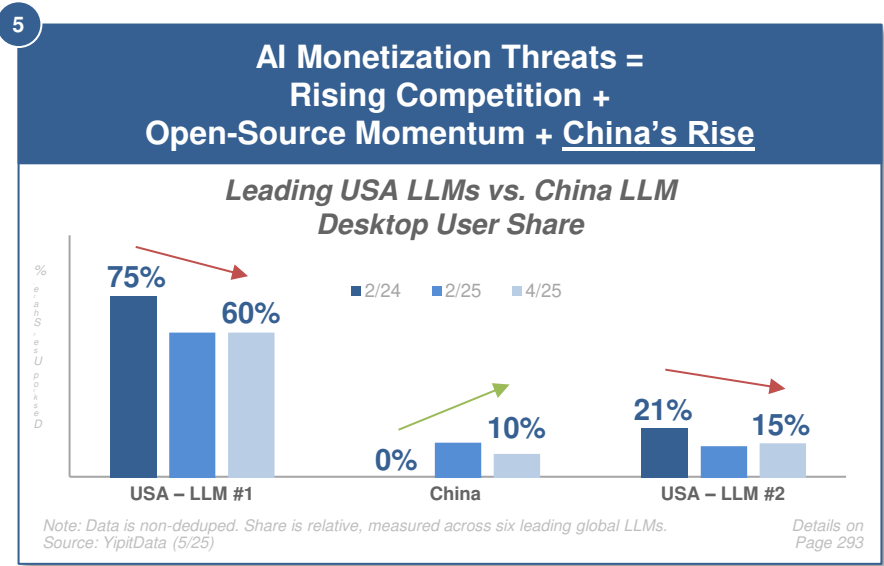
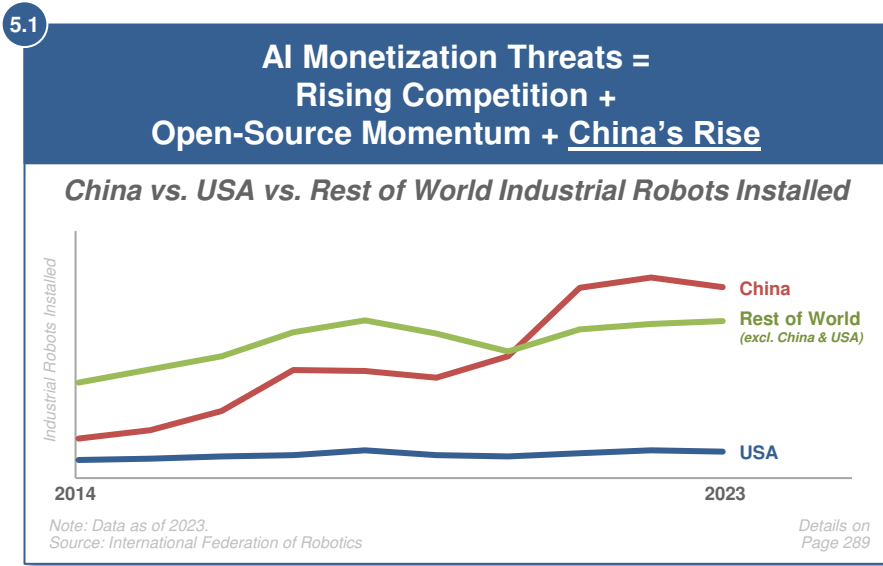
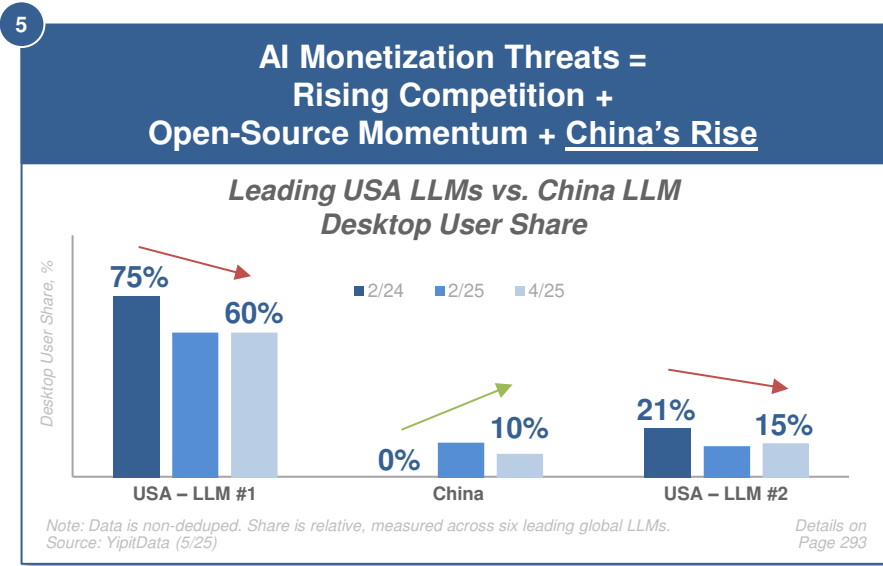
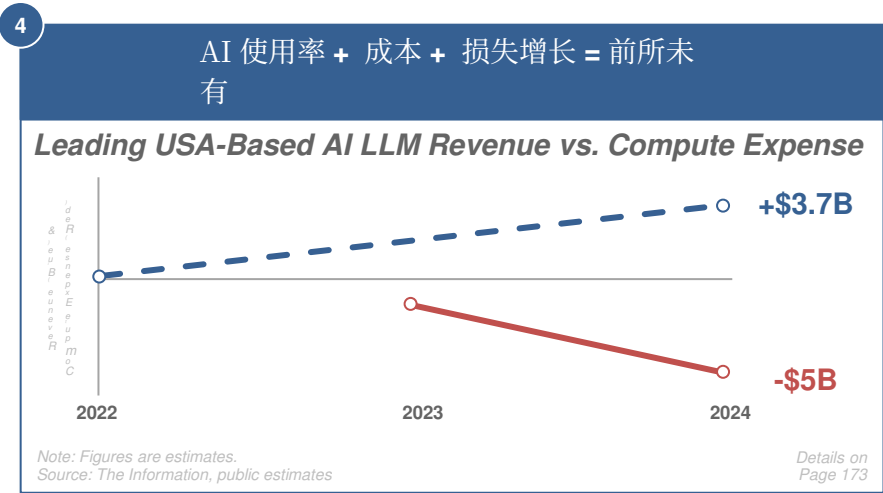
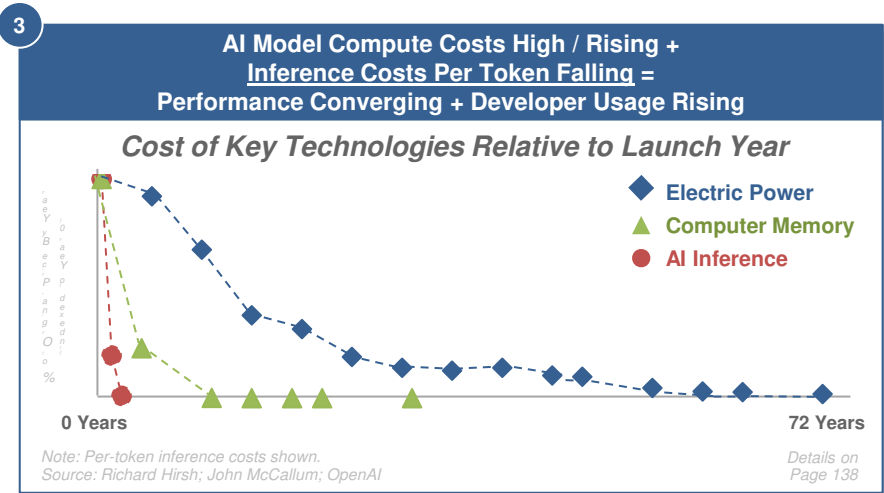
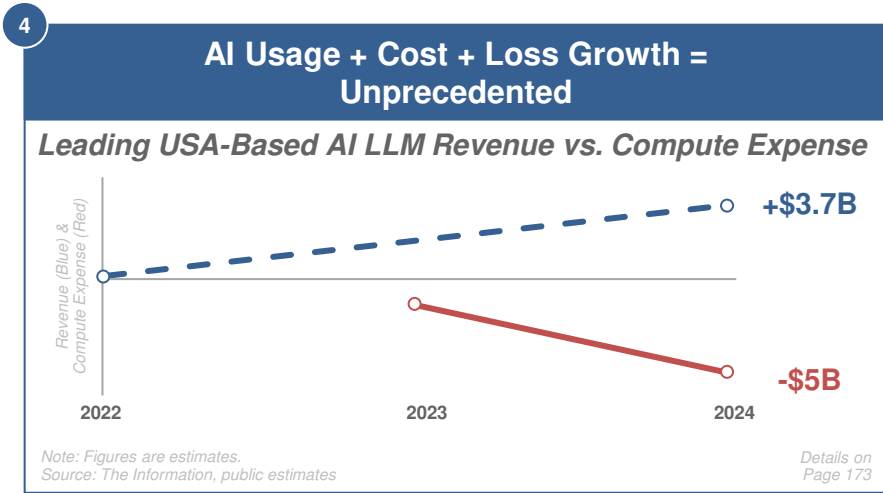
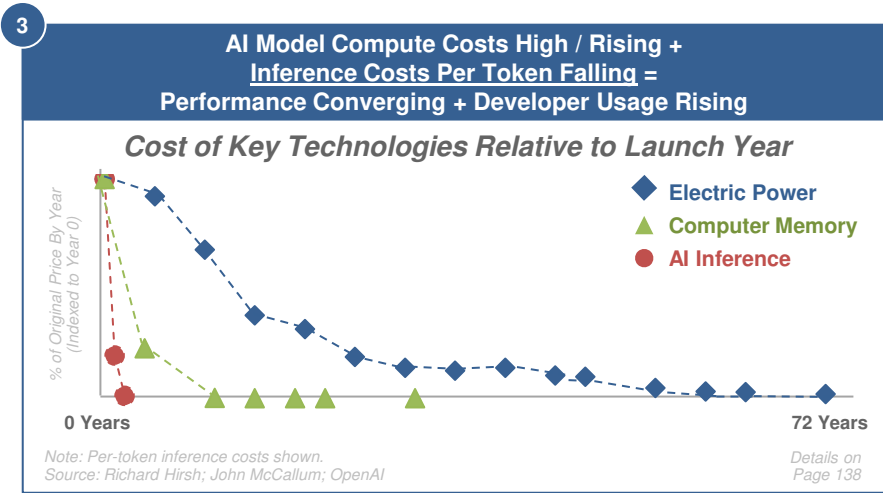
Special thanks to Grant Watson and Keeyan Sanjasaz and BOND colleagues who helped steer ideas and bring this report to life. And, to the many friends and technology builders who helped, directly or via your work, and are driving technology forward.

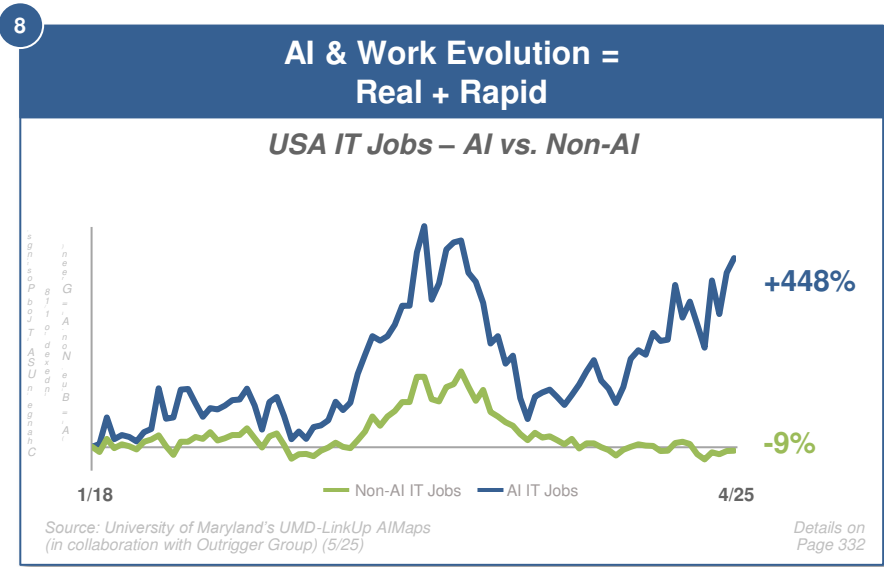
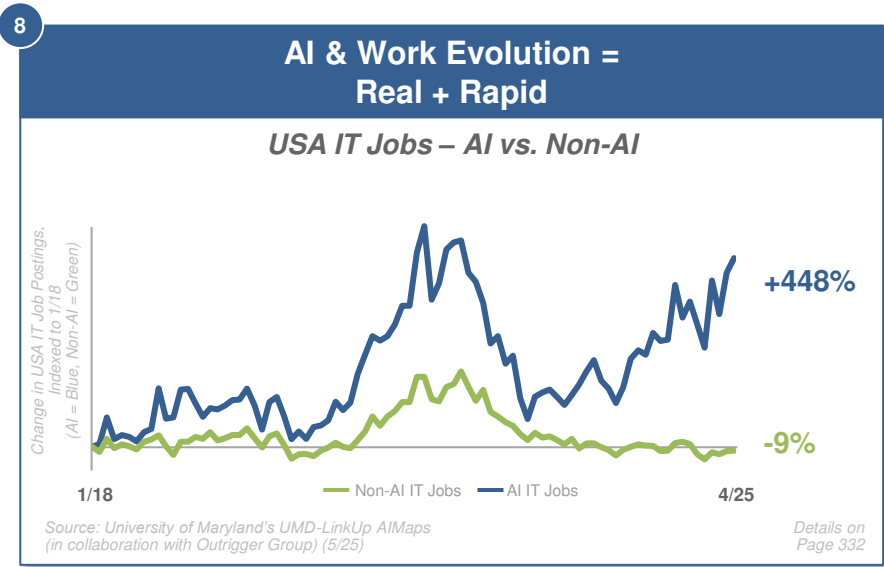
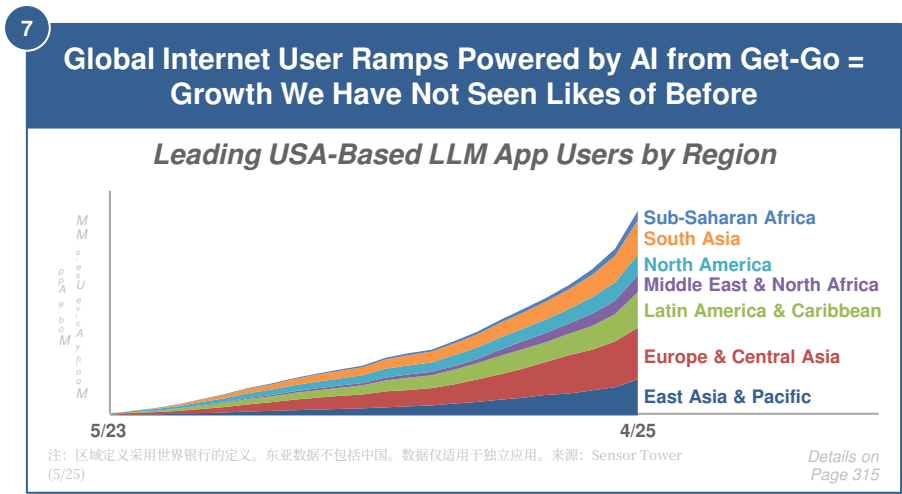
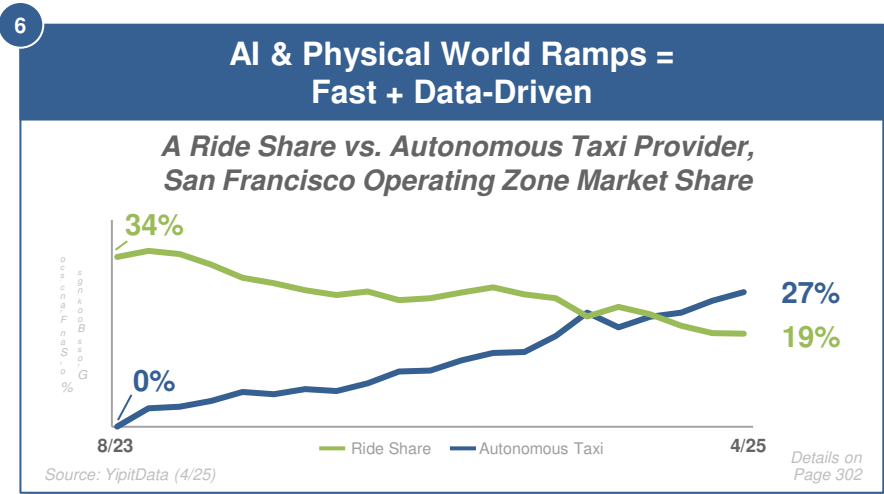
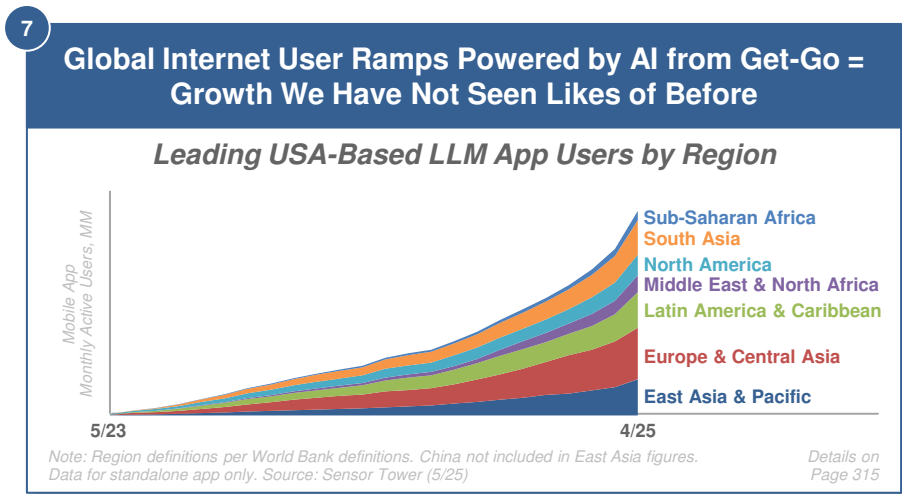
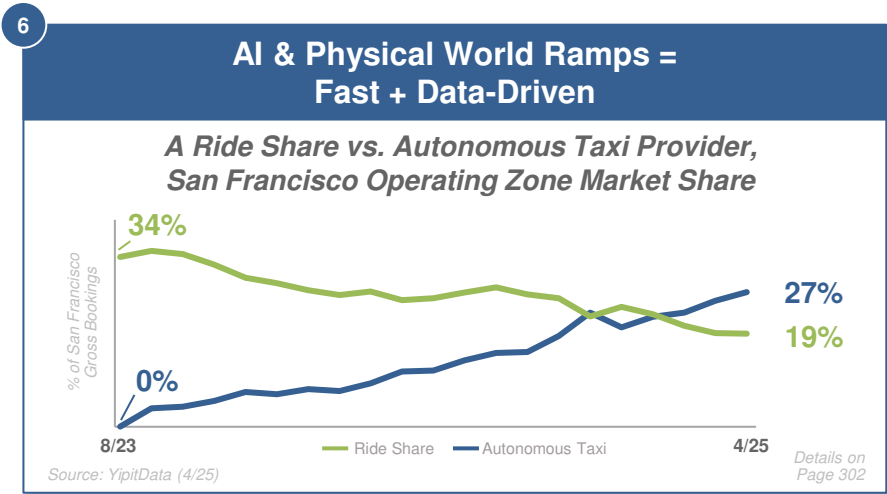
	#
1 <b>Seem Like Change Happening Faster Than Ever?</b> Yes, It Is	9-51
2 <b>AI User + Usage + CapEx Growth =</b> Unprecedented	52-128
3 <b>AI Model Compute Costs High / Rising + Inference Costs Per Token Falling =</b> Performance Converging + Developer Usage Rising	129-152
4 <b>AI Usage + Cost + Loss Growth =</b> Unprecedented	153-247
5 <b>AI Monetization Threats =</b> Rising Competition + Open-Source Momentum + China’s Rise	248-298
6 <b>AI &amp; Physical World Ramps =</b> Fast + Data-Driven	299-307
7 <b>Global Internet User Ramps Powered by AI from Get-Go =</b> Growth We Have Not Seen Likes of Before	308-322
8 <b>AI &amp; Work Evolution =</b> Real + Rapid	323-336
↓	

	#
1 变化似乎比以往任何时候都快？是的，确实如此	9-51
2 AI 用户 + 使用情况 + 资本支出增长 = 前所 未有	52-128
3 AI 模型计算成本高 / 上升 + 每次 Token 的推理成本下降 = 性能趋同 + 开发者使用量上升	129-152
4 AI 使用情况 + 成本 + 损失增长 = 前所未 有	153-247
5 人工智能货币化威胁 = 日益激烈的竞争 + 开源势头 + 中国的崛起	248-298
6 AI 与物理世界的发展 = 快速 + 数据 驱动	299-307
7 <b>Global Internet User Ramps Powered by AI from Get-Go =</b> 我们以前 从未见过的增长	308-322
8 AI 与工作演变 = 真实 + 迅 速	323-336
↓	









To say the world is changing at unprecedented rates is an understatement.  
Rapid and transformative technology innovation / adoption represent key underpinnings of these changes.  
As does leadership evolution for the global powers.

Google’s founding mission (1998) was to ‘organize the world’s information and make it universally accessible and useful.’  
Alibaba’s founding mission (1999) was to ‘make it easy to do business anywhere.’  
Facebook’s founding mission (2004) was ‘to give people the power to share and make the world more open and connected.’

Fast forward to today with the world’s organized, connected and accessible information being supercharged by artificial intelligence, accelerating computing power, and semi-borderless capital...all driving massive change.

Sport provides a good analogy for AI’s constant improvements. As athletes continue to wow us and break records, their talent is increasingly enhanced by better data / inputs / training.  
The same is true for businesses, where computers are ingesting massive datasets to get smarter and more competitive. Breakthroughs in large models, cost-per-token declines, open-source proliferation and chip performance improvements are making new tech advances increasingly more powerful, accessible, and economically viable.

OpenAI’s ChatGPT – based on user / usage / monetization metrics – is history’s biggest ‘overnight’ success (nine years post-founding). AI usage is surging among consumers, developers, enterprises and governments.  
And unlike the Internet 1.0 revolution – where technology started in the USA and steadily diffused globally – ChatGPT hit the world stage all at once, growing in most global regions simultaneously.

Meanwhile, platform incumbents and emerging challengers are racing to build and deploy the next layers of AI infrastructure: agentic interfaces, enterprise copilots, real-world autonomous systems, and sovereign models.

Rapid advances in artificial intelligence, compute infrastructure, and global connectivity are fundamentally reshaping how work gets done, how capital is deployed, and how leadership is defined – across both companies and countries.

At the same time, we have leadership evolution among the global powers, each of whom is challenging the other’s competitive and comparative advantage. We see the world’s most powerful countries revved up by varying degrees of economic / societal / territorial aspiration...

To say the world is changing at unprecedented rates is an understatement.  
Rapid and transformative technology innovation / adoption represent key underpinnings of these changes.  
全球大国的领导力演变也是如此。

谷歌的创立使命（1998 年）是 “ 整理世界信息，使其人人皆可访问并从中受益 ”。阿里巴巴的创立使命（1999 年）是 “ 让天下没有难做的生意 ”。Facebook 的创立使命（2004 年）是 “ 赋予人们分享的力量，让世界更开放、更互联 ”。

快进到今天，随着世界有组织的、互联的和可访问的信息被人工智能、加速计算能力和半无边界资本所增压 …… 所有这些都在推动巨大的变革。

体育为人工智能的不断改进提供了一个很好的类比。随着运动员不断让我们惊叹并打破纪录，他们的天赋越来越受到更好的数据 / 投入 / 训练的加强。对于企业来说也是如此，计算机正在摄取大量数据集，以变得更智能、更具竞争力。大型模型的突破、每 token 成本的下降、开源的普及和芯片性能的提高，使得新的技术进步越来越强大、可访问且在经济上可行。

OpenAI 的 ChatGPT – 基于用户 / 使用情况 / 货币化指标 – 是历史上最大的 “ 一夜 ” 成功（成立九年后）。消费者、开发者、企业和政府对人工智能的使用正在激增。与互联网 1.0 革命不同 – 技术起源于美国并稳步向全球扩散 – ChatGPT 一下子风靡全球，并在全球大多数地区同时发展。

与此同时，平台巨头和新兴挑战者正在竞相构建和部署下一层人工智能基础设施：代理接口、企业副驾驶、现实世界的自主系统和主权模型。

人工智能、计算基础设施和全球连接的快速发展正在从根本上改变工作完成的方式、资本部署的方式以及领导力的定义 – 无论是在公司还是国家层面。

与此同时，我们看到全球大国之间的领导力演变，每个国家都在挑战其他国家的竞争优势和比较优势。我们看到世界上最强大的国家受到不同程度的经济 / 社会 / 领土愿望的激励 ……

...Increasingly, two hefty forces – technological and geopolitical – are intertwining. Andrew Bosworth (Meta Platforms CTO), on a recent ‘Possible’ podcast described the current state of AI as *our space race and the people we’re discussing, especially China, are highly capable... there’s very few secrets. And there’s just progress. And you want to make sure that you’re never behind.*

The reality is AI leadership could beget geopolitical leadership – and not vice-versa.

This state of affairs brings tremendous uncertainty...yet it leads us back to one of our favorite quotes – *Statistically speaking, the world doesn’t end that often*, from former T. Rowe Price Chairman and CEO Brian Rogers.

As investors, we always assume everything can go wrong, but the exciting part is the consideration of what can go right. Time and time again, the case for optimism is one of the best bets one can make. The magic of watching AI do your work for you feels like the early days of email and web search – technologies that fundamentally changed our world. The better / faster / cheaper impacts of AI seem just as magical, but even quicker.

No doubt, these are also dangerous and uncertain times. But a long-term case for optimism for artificial intelligence is based on the idea that intense competition and innovation... increasingly-accessible compute...rapidly-rising global adoption of AI-infused technology...and thoughtful and calculated leadership can foster sufficient trepidation and respect, that in turn, could lead to Mutually Assured Deterrence.

For some, the evolution of AI will create a race to the bottom; for others, it will create a race to the top. The speculative and frenetic forces of capitalism and creative destruction are tectonic. It’s undeniable that it’s ‘game on,’ especially with the USA and China and the tech powerhouses charging ahead.

In this document, we share data / research / benchmarks from third parties that use methodologies they deem to be effective – we are thankful for the hard work so many are doing to illustrate trending during this uniquely dynamic time. Our goal is to add to the discussion.

... 日益增强的是，两种强大的力量 —— 技术和地缘政治 —— 正在交织。Meta Platforms CTO Andrew Bosworth 在最近的 “Possible” 播客中将当前的人工智能状态描述为我们的太空竞赛，而且我们正在讨论的人，特别是中国，非常有能力 …… 几乎没有什么秘密。而且一直在进步。你想要确保你永远不会落后。

现实情况是，人工智能的领导地位可能会带来地缘政治的领导地位 —— 而不是相反。

这种状态带来了巨大的不确定性 …… 但它又让我们回想起我们最喜欢的一句话 —— 从统计学上讲，世界不会经常终结，这句话来自前 T. Rowe Price 董事长兼首席执行官 Brian Rogers。

作为投资者，我们总是假设一切都可能出错，但令人兴奋的部分是考虑什么可以做对。一次又一次，乐观的理由是人们可以做的最好的赌注之一。看着人工智能为你工作的感觉就像电子邮件和网络搜索的早期 —— 从根本上改变了我们世界的技术。人工智能更好 / 更快 / 更便宜的影响似乎同样神奇，但速度更快。

毫无疑问，现在也是一个充满危险和不确定的时代。但对人工智能持乐观态度的长期理由是，激烈的竞争和创新 … 日益普及的计算 … 迅速上升的全球对人工智能技术的采用 … 以及周到且经过计算的领导力能够培养足够的恐惧和尊重，进而可能导致相互确保威慑。

对一些人来说，人工智能的进化将导致逐底竞争；对另一些人来说，它将导致逐顶竞争。资本主义和创造性破坏的投机性和疯狂力量是巨大的。不可否认的是，现在是 “游戏开始” 了，尤其是美国、中国和科技巨头都在向前冲。

在本文件中，我们分享来自第三方的使用他们认为有效的方法论的数据 / 研究 / 基准 —— 我们感谢如此多的人为说明这个独特动态时期的趋势所做的辛勤工作。我们的目标是为讨论添加内容。

- 1 **Seem Like Change Happening Faster Than Ever?**  
Yes, It Is
- 2 **AI User + Usage + CapEx Growth =**  
Unprecedented
- 3 **AI Model Compute Costs High / Rising + Inference Costs Per Token Falling =**  
Performance Converging + Developer Usage Rising
- 4 **AI Usage + Cost + Loss Growth =**  
Unprecedented
- 5 **AI Monetization Threats =**  
Rising Competition + Open-Source Momentum + China's Rise
- 6 **AI & Physical World Ramps =**  
Fast + Data-Driven
- 7 **Global Internet User Ramps Powered by AI from Get-Go =**  
Growth We Have Not Seen Likes of Before
- 8 **AI & Work Evolution =**  
Real + Rapid

- 1 变化似乎比以往任何时候都快？是的，确实如此
- 2 AI 用户 + 使用量 + 资本支出增长 = 前所未有
- 3 AI 模型计算成本高 / 上涨 + 每次 Token 的推理成本下降 = 性能趋同 + 开发者使用量上升
- 4 AI 使用量 + 成本 + 损失增长 = 前所未有
- 5 AI 货币化威胁 = 日益激烈的竞争 + 开源势头 + 中国的崛起
- 6 AI 和物理世界发展 = 快速 + 数据驱动
- 7 **Global Internet User Ramps Powered by AI from Get-Go =** 我们从未见过的增长
- 8 AI 与工作演变 = 真实 + 迅速

Technology Compounding =  
Numbers Behind The Momentum

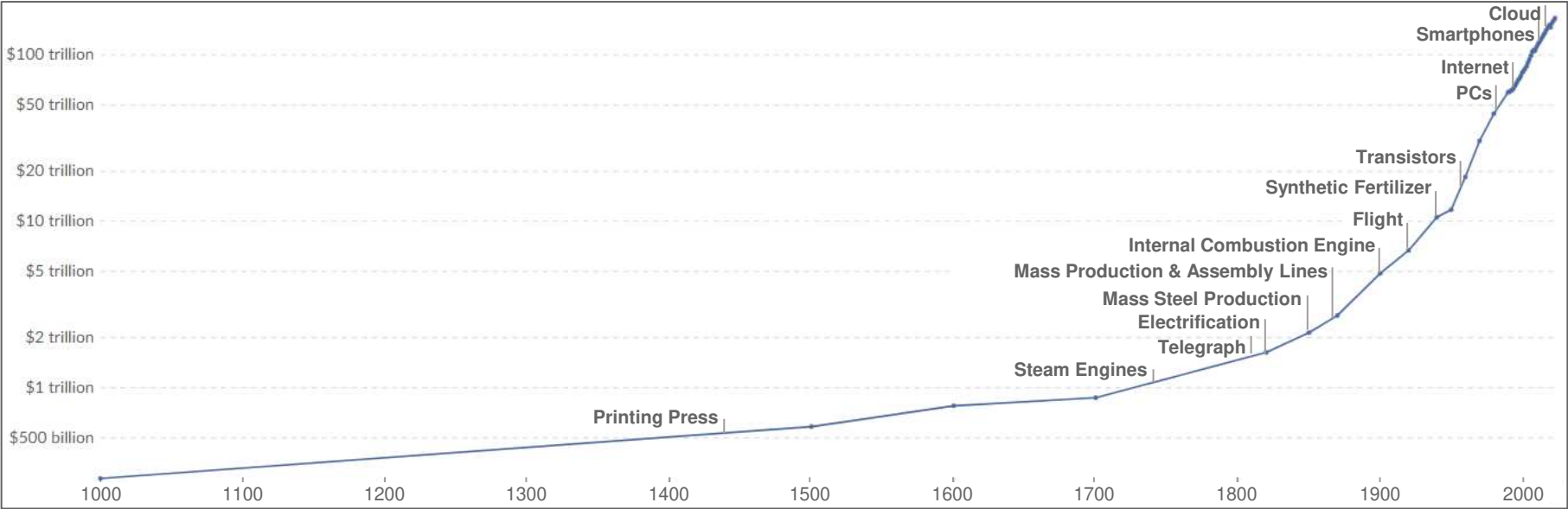
技术复合 =  
Numbers Behind The Momentum



Technology Compounding Over Thousand-Plus Years =  
Better + Faster + Cheaper → More...

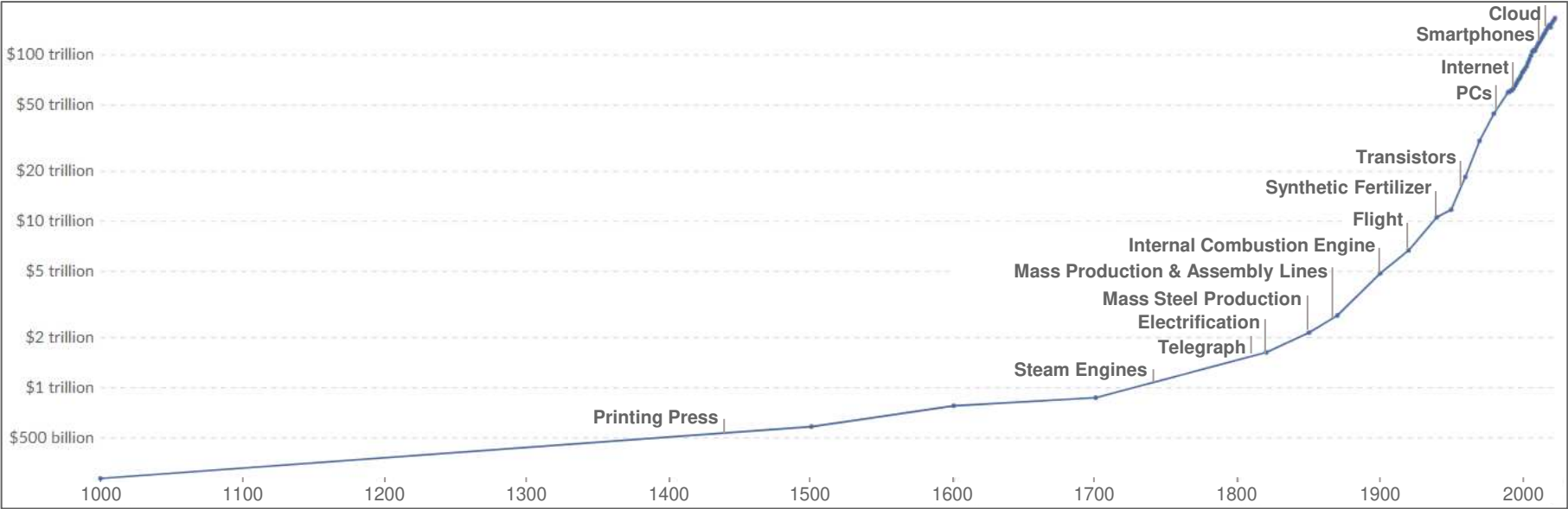
技术在千年以上的积累 = 更好 + 更快 + 更便宜 → 更多 ……

Global GDP – Last 1,000+ Years, per Maddison Project



*Note: Chart expressed in trillions of real GDP as measured by 2011 'GK\$' on a logarithmic scale. GK\$ (Gross Knowledge Dollars) is an informal term used to estimate the potential business value of a specific insight, idea, or proprietary knowledge. It reflects how much that knowledge could be worth if applied effectively, even if it hasn't yet generated revenue. Source: Microsoft, 'Governing AI: A Blueprint for the Future,' Microsoft Report (5/23); Data via Maddison Project & Our World in Data*

全球 GDP – 过去 1,000+ 年，来自 Maddison 项目

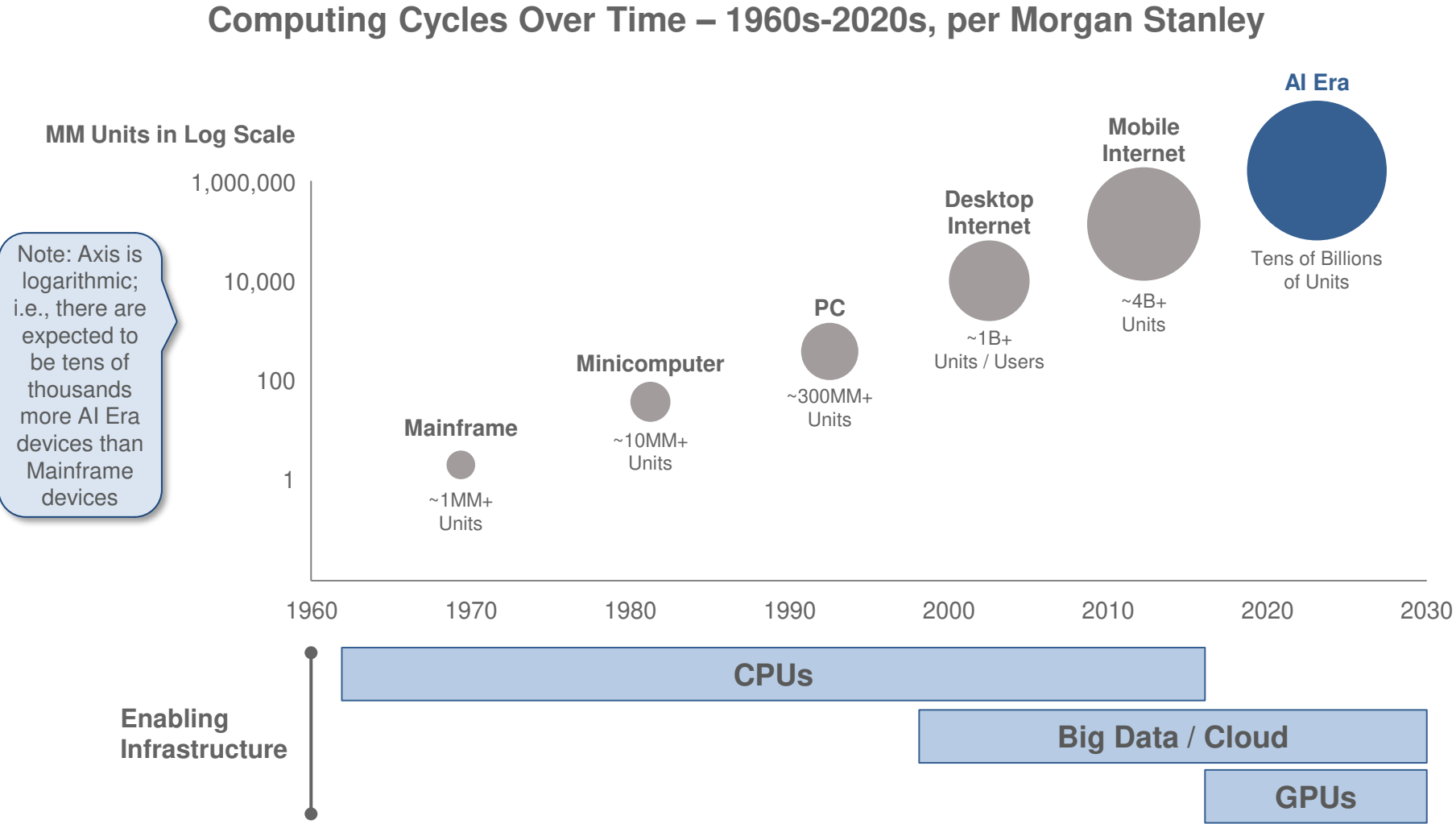


注：图表以万亿实际 GDP 表示，以 2011 年的“GK\$”为单位，采用对数比例。GK\$（知识总额美元）是一个非正式术语，用于估计特定见解、想法或专有知识的潜在商业价值。它反映了即使该知识尚未产生收入，如果有效应用，它可能值多少钱。来源：微软，“人工智能治理：未来蓝图”，微软报告（5/23）；数据来自 Maddison 项目和 Our World in Data

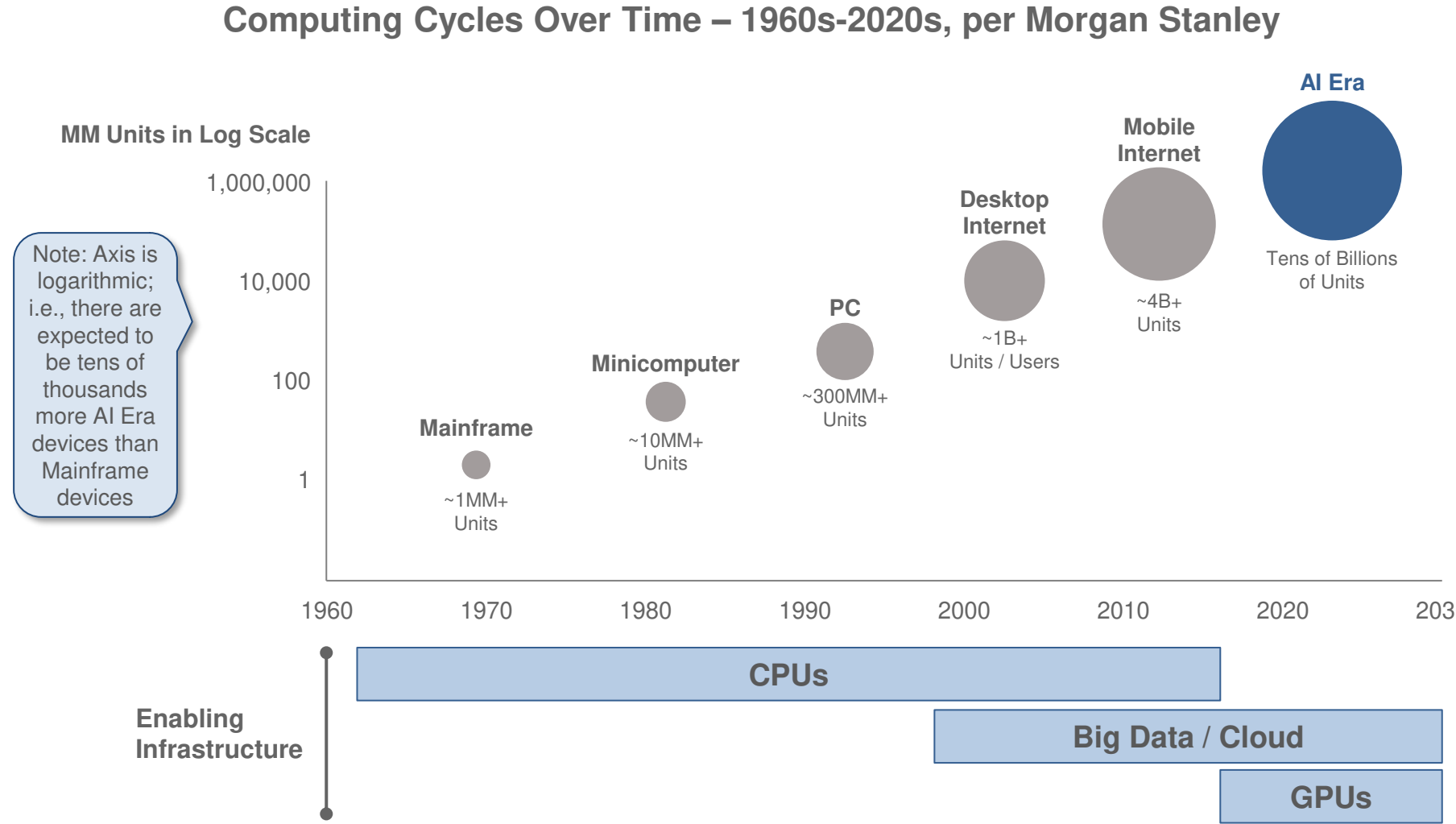


...Technology Compounding Over Fifty-Plus Years = Better + Faster + Cheaper → More

… 技术在超过五十年的时间里不断积累 = 更好 + 更快 + 更便宜 → 更多



Note: PC units as of 2000. Desktop internet users as of 2005, installed base as of 2010. Mobile internet units are the installed based of smartphones & tablets in 2020. Cloud & data center capex includes Google, Amazon, Microsoft, Meta, Alibaba, Apple, IBM, Oracle, Tencent, & Baidu for ten years ending 2022. 'Tens of billions of units' refers to the potential device & user base that could end up using AI technology; this includes smartphones, IOT devices, robotics, etc. Source: Weiss et al. 'AI Index: Mapping the \$4 Trillion Enterprise Impact' via Morgan Stanley (10/23)



注：PC 单位截至 2000 年。桌面互联网用户截至 2005 年，安装基数截至 2010 年。移动互联网单位是 2020 年智能手机和平板电脑的安装基数。云和数据中心资本支出包括谷歌、亚马逊、微软、Meta、阿里巴巴、苹果、IBM、甲骨文、腾讯和百度截至 2022 年的十年。“数百亿台设备”指的是可能最终使用人工智能技术的潜在设备 & 用户群；这包括智能手机、物联网设备、机器人等。资料来源：Weiss et al. 通过以下途径发布的“人工智能指数：绘制 4 万亿美元的企业影响图”摩根士丹利（10/23）

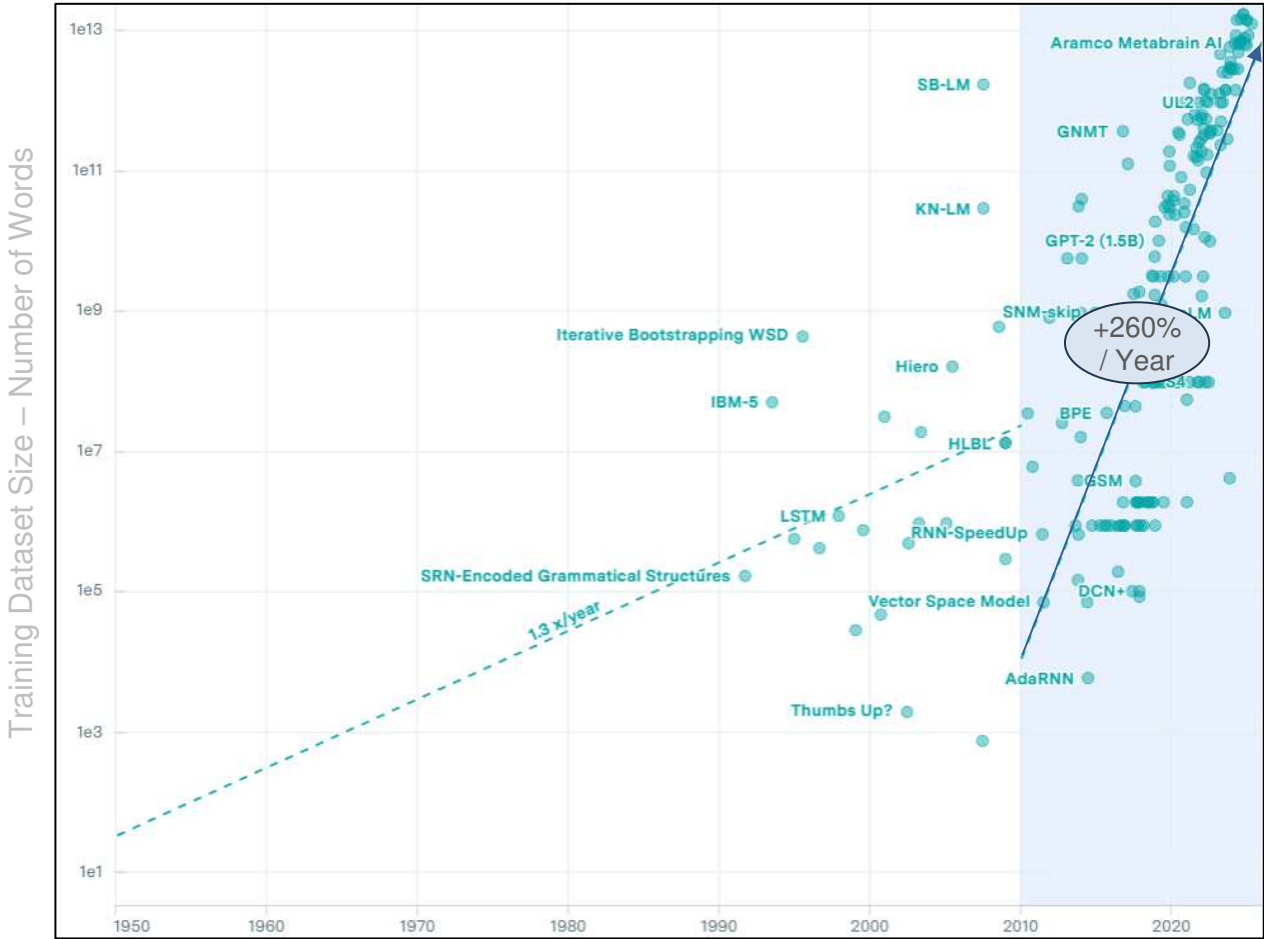
AI Technology Compounding =  
Numbers Behind The Momentum

AI Technology Compounding =  
Numbers Behind The Momentum

260% Annual Growth Over Fifteen Years of...  
Data to Train AI Models Led To...

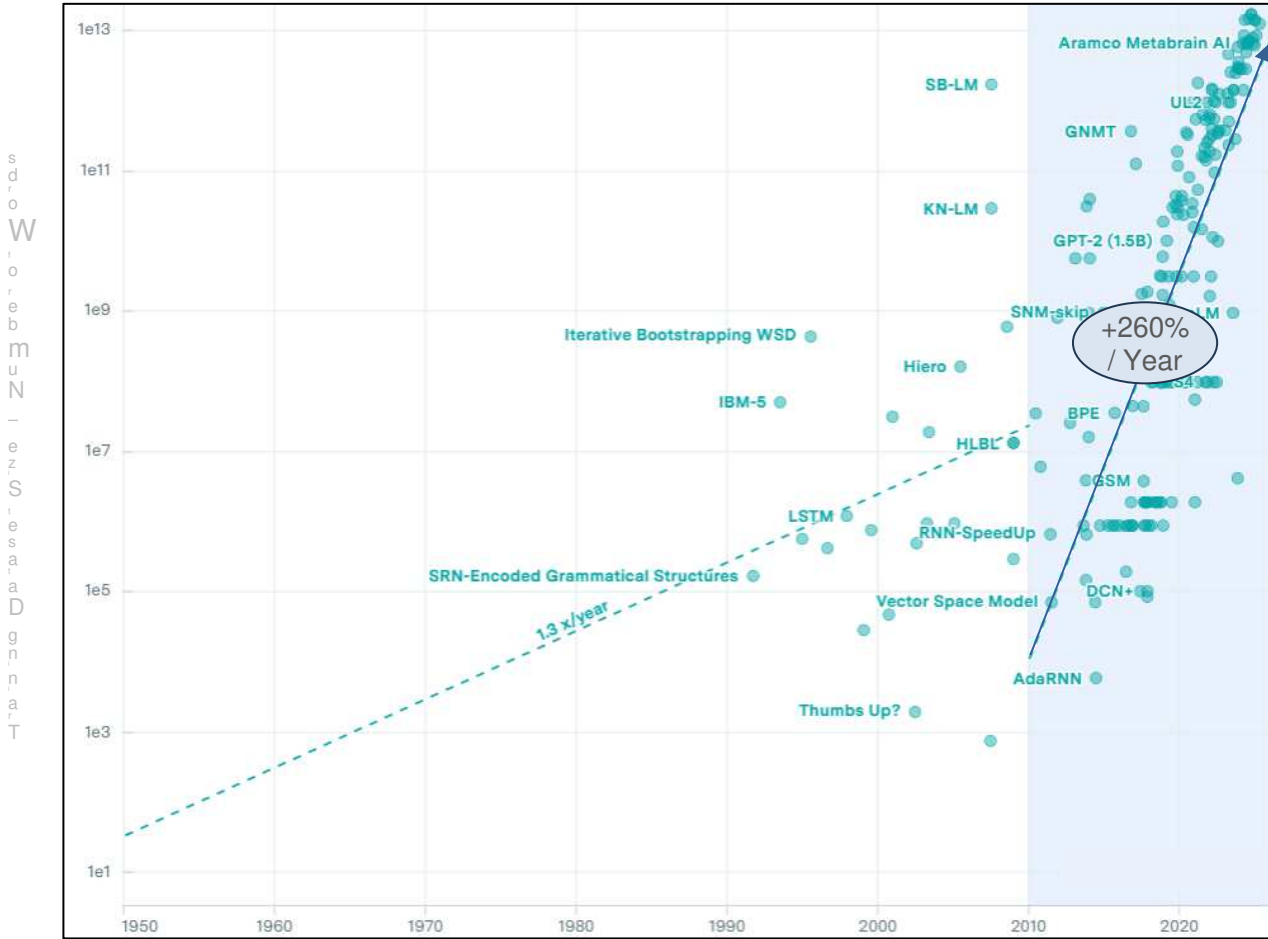
在 15 年的时间里，年增长率达到 260%…… 用于  
训练人工智能模型的数据导致了 ……

Training Dataset Size (Number of Words) for Key AI Models – 1950-2025, per Epoch AI



Note: Only “notable” language models shown (per Epoch AI, includes state of the art improvement on a recognized benchmark, >1K citations, historically relevant, with significant use).  
Source: Epoch AI (5/25)

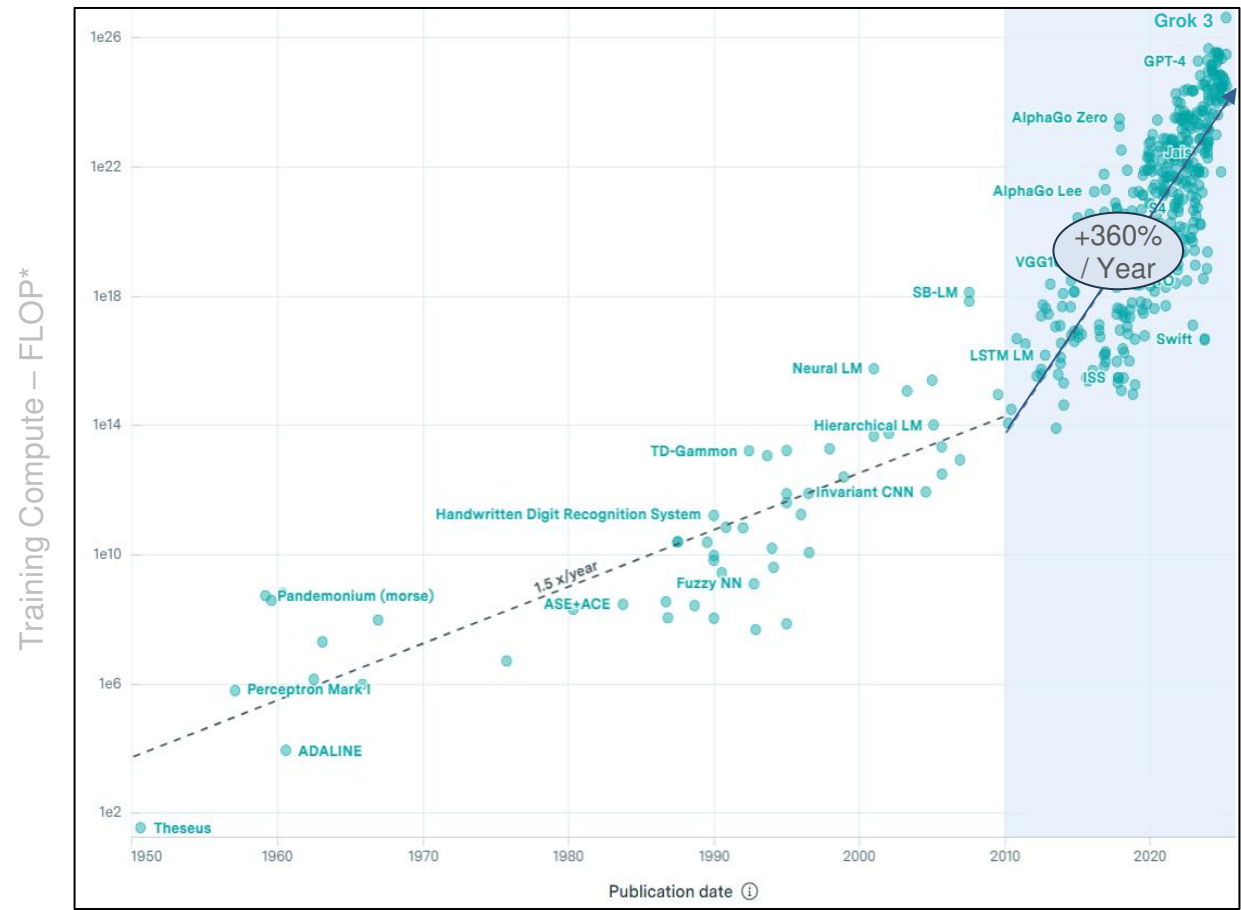
Training Dataset Size (Number of Words) for Key AI Models – 1950-2025, per Epoch AI



注意：仅显示“值得注意的”语言模型（根据 Epoch AI，包括对公认基准的最新改进，>1K 引用，具有历史相关性，并具有重要用途）。来源：Epoch AI (5/25)

...360% Annual Growth Over Fifteen Years of...  
Compute to Train AI Models Led To...

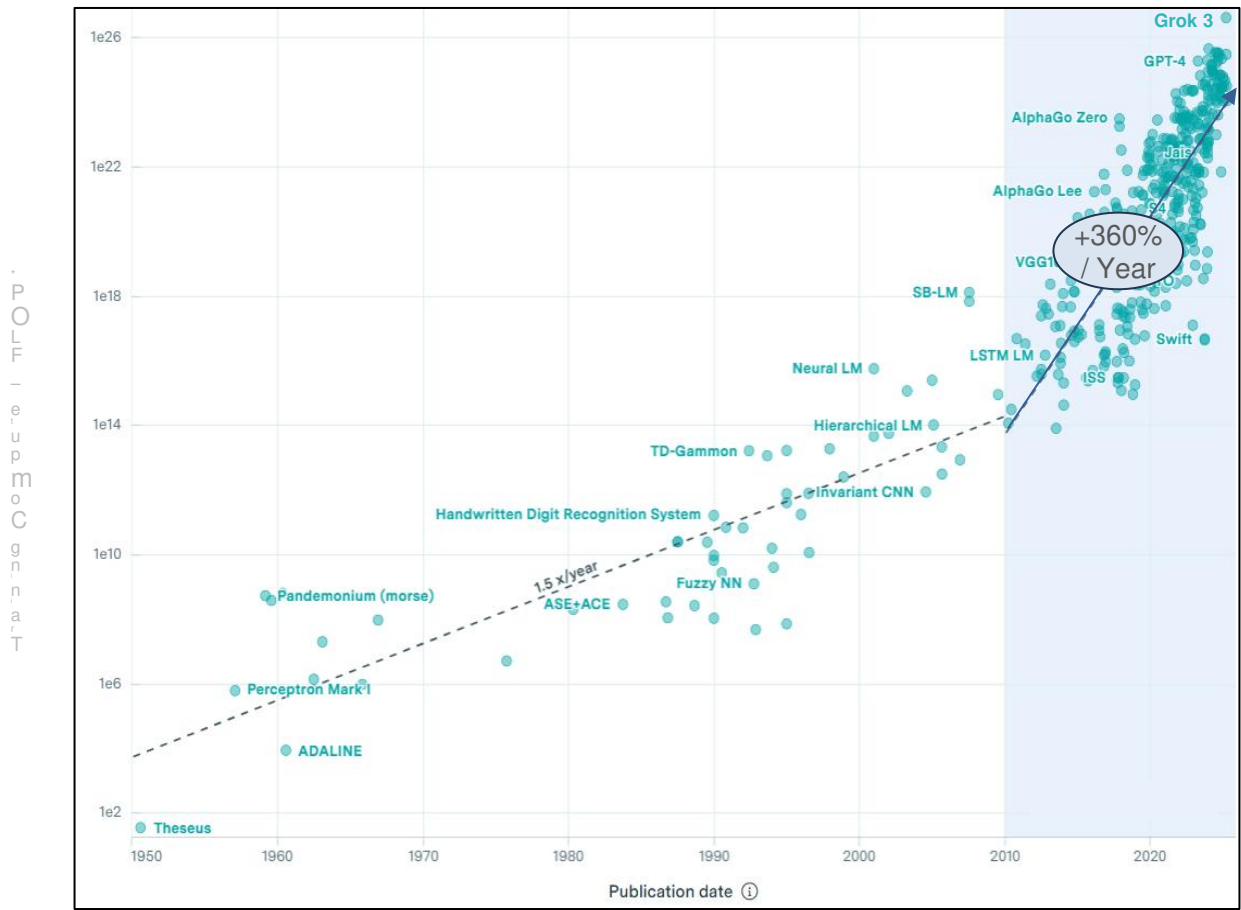
Training Compute (FLOP) for Key AI Models – 1950-2025, per Epoch AI



*\*A FLOP (floating point operation) is a basic unit of computation used to measure processing power, representing a single arithmetic calculation involving decimal numbers. In AI, total FLOPs are often used to estimate the computational cost of training or running a model.  
Note: Only language models shown (per Epoch AI, includes state of the art improvement on a recognized benchmark, >1K citations, historically relevant, with significant use). Source: Epoch AI (5/25)*

...360% Annual Growth Over Fifteen Years of...  
Compute to Train AI Models Led To...

Training Compute (FLOP) for Key AI Models – 1950-2025, per Epoch AI

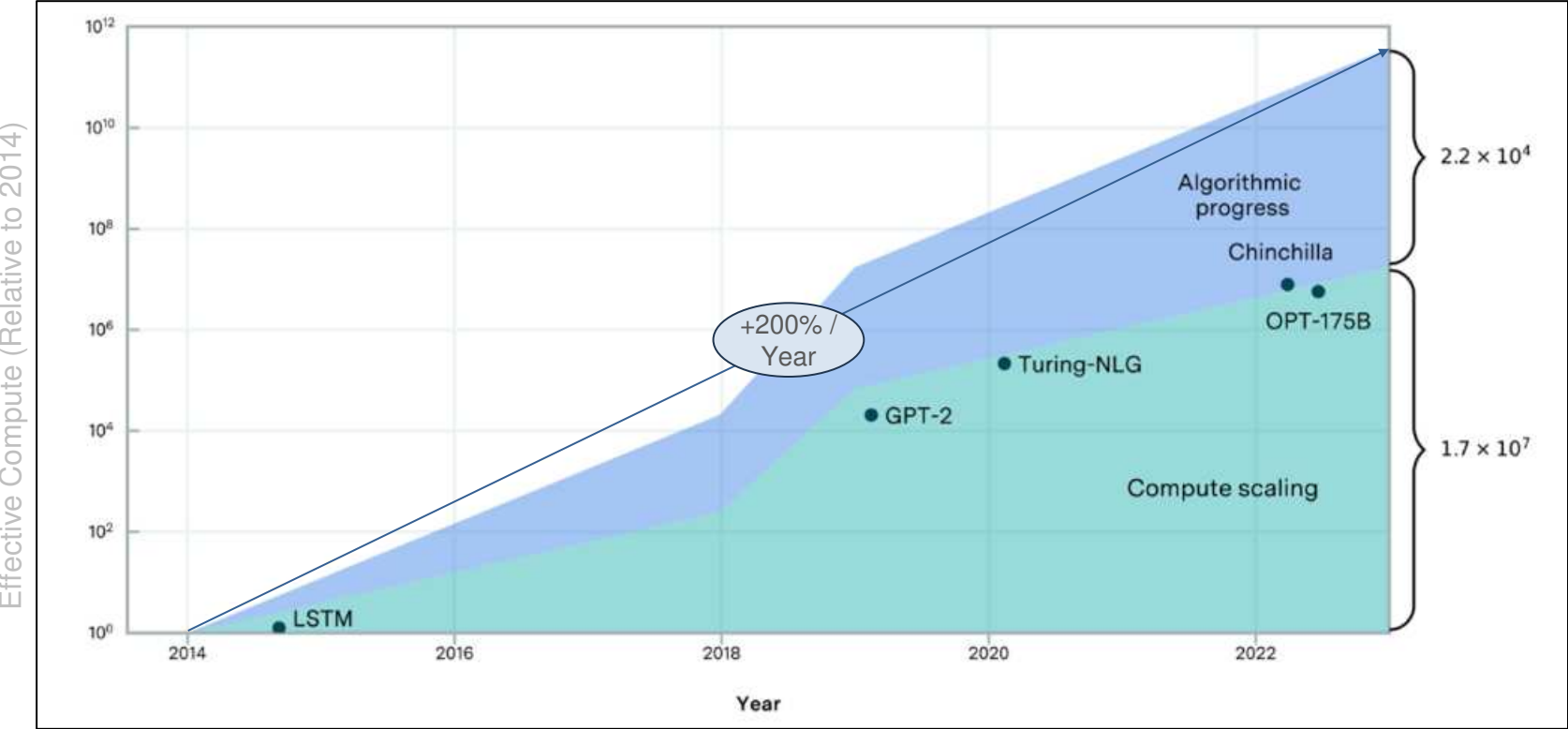


*\*FLOP（浮点运算）是用于衡量处理能力的基本计算单位，表示涉及十进制数的单个算术计算。在人工智能中，总 FLOP 通常用于估计训练或运行模型的计算成本。注意：仅显示语言模型（根据 Epoch AI，包括对公认基准的最新改进，>1K 引用，具有历史意义，且具有重要用途）。来源：Epoch AI (5/25)*

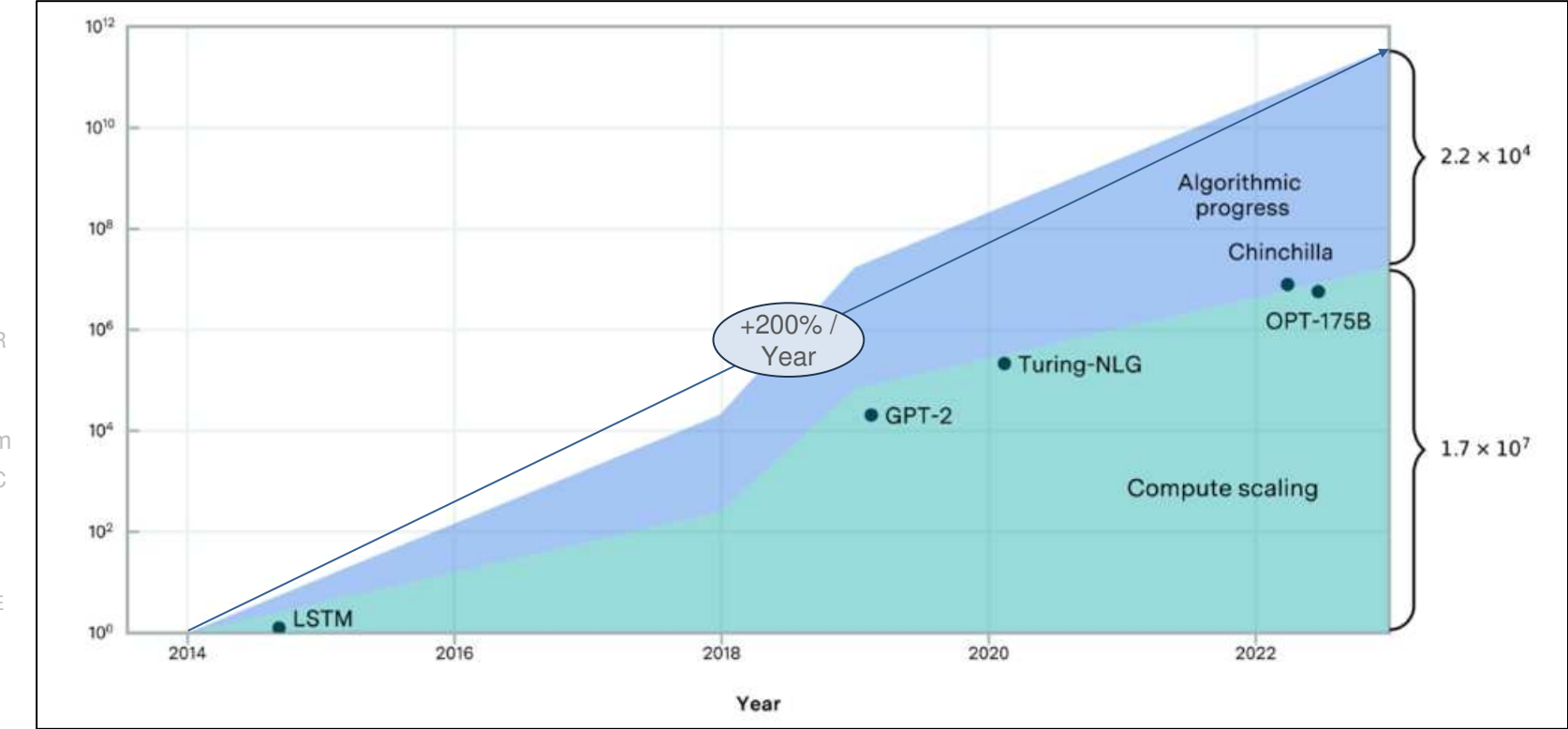
...200% Annual Growth Over Nine Years of...  
Compute Gains from Better Algorithms Led To...

…200% 的九年年增长率 … 更好的算法带来的计算收益 …

Impact of Improved Algorithms on AI Model Performance – 2014-2023, per Epoch AI



Impact of Improved Algorithms on AI Model Performance – 2014-2023, per Epoch AI



Note: Estimates how much progress comes from bigger models versus smarter algorithms, based on how much computing power you'd need to reach top performance without any improvements. Source: Epoch AI (3/24)

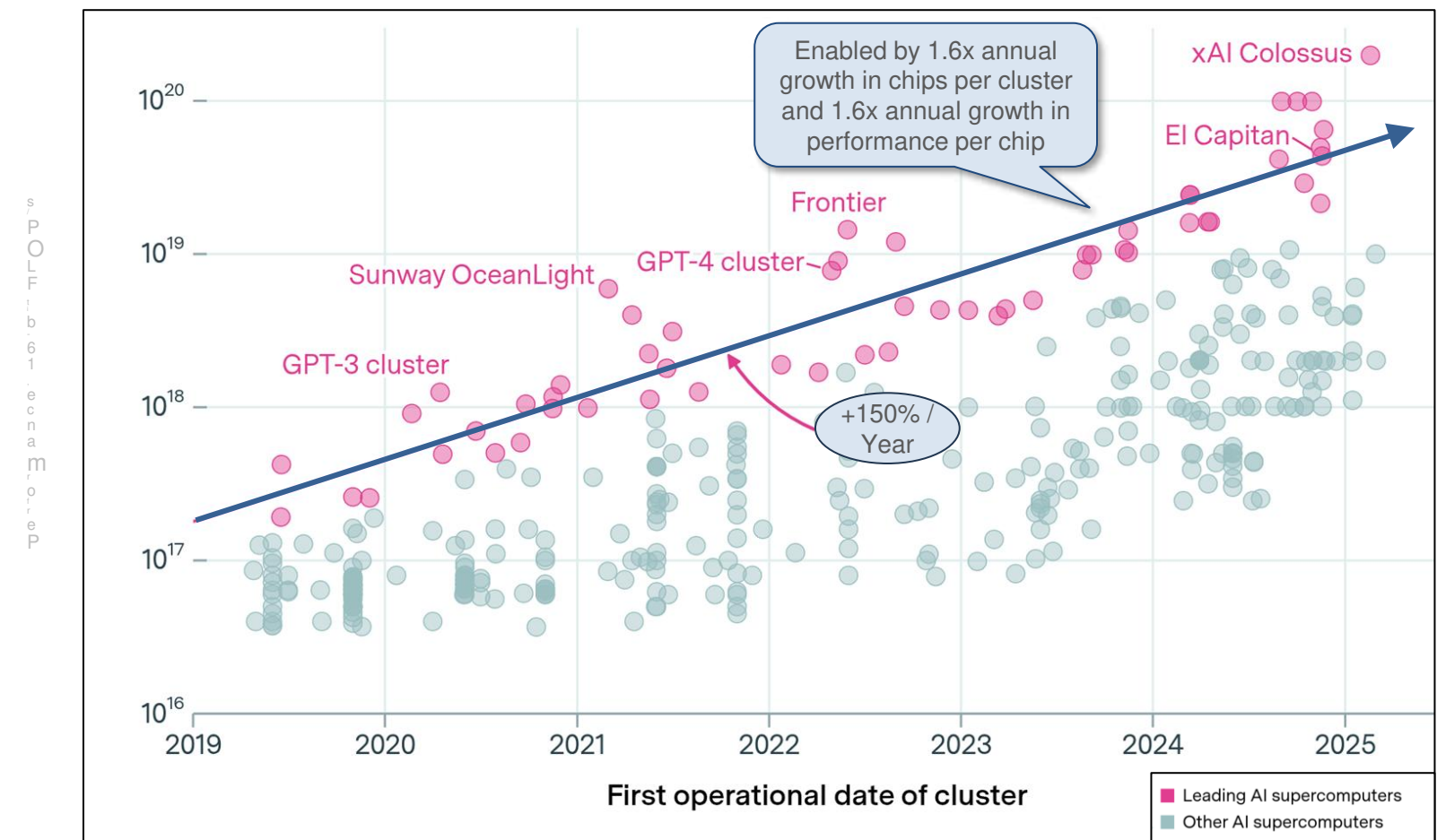
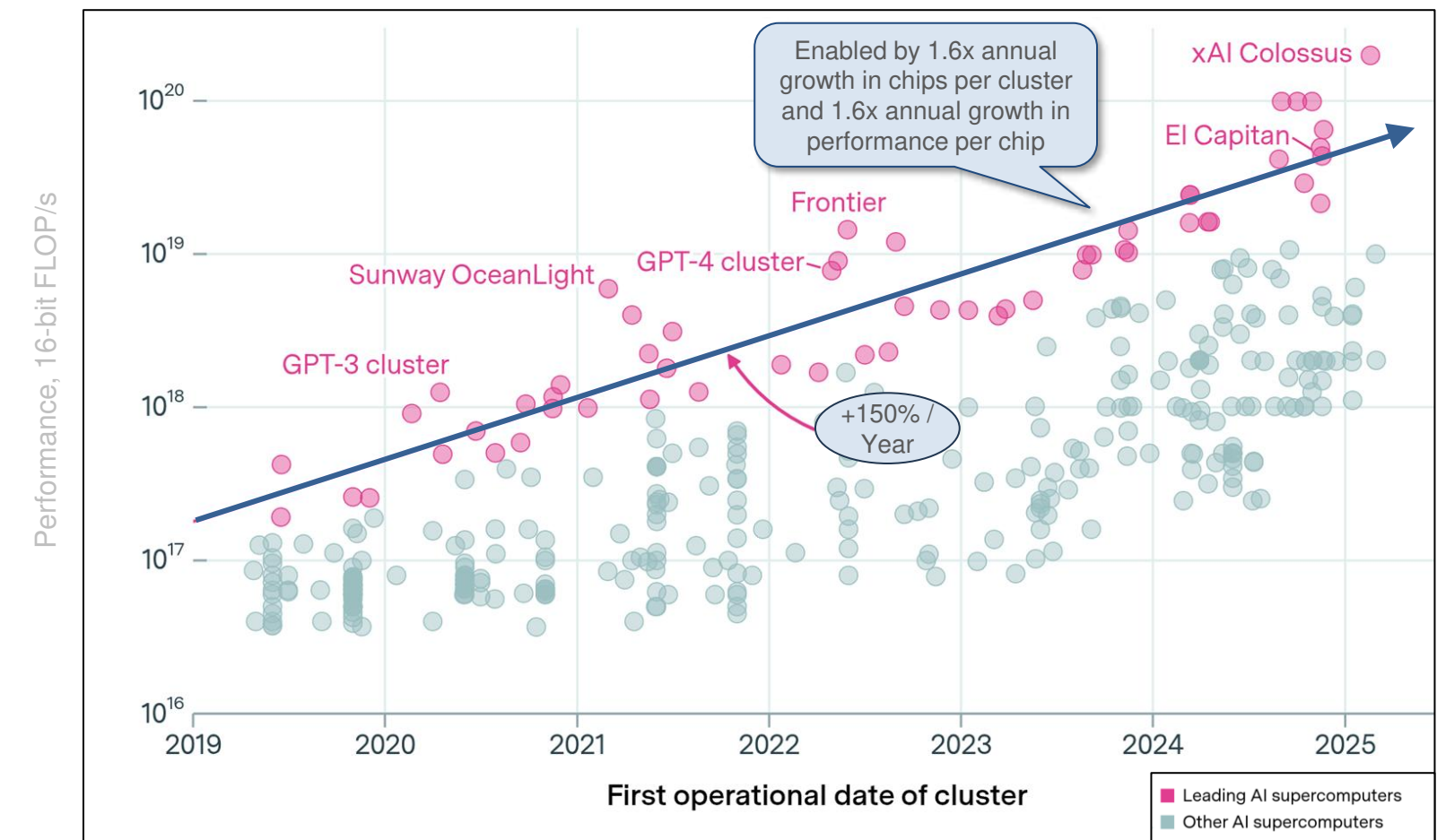
注意：根据在没有任何改进的情况下达到最佳性能所需的计算能力，估算有多少进展来自更大的模型与更智能的算法。来源：Epoch AI (3/24)

...150% Annual Growth Over Six Years of...  
Performance Gains from Better AI Supercomputers Led To...

… 六年内年增长率达 150%，更优的 AI 超级计算机带来了性能提升  
…

Performance of Leading AI Supercomputers (FLOP/s) – 2019-2025, per Epoch AI

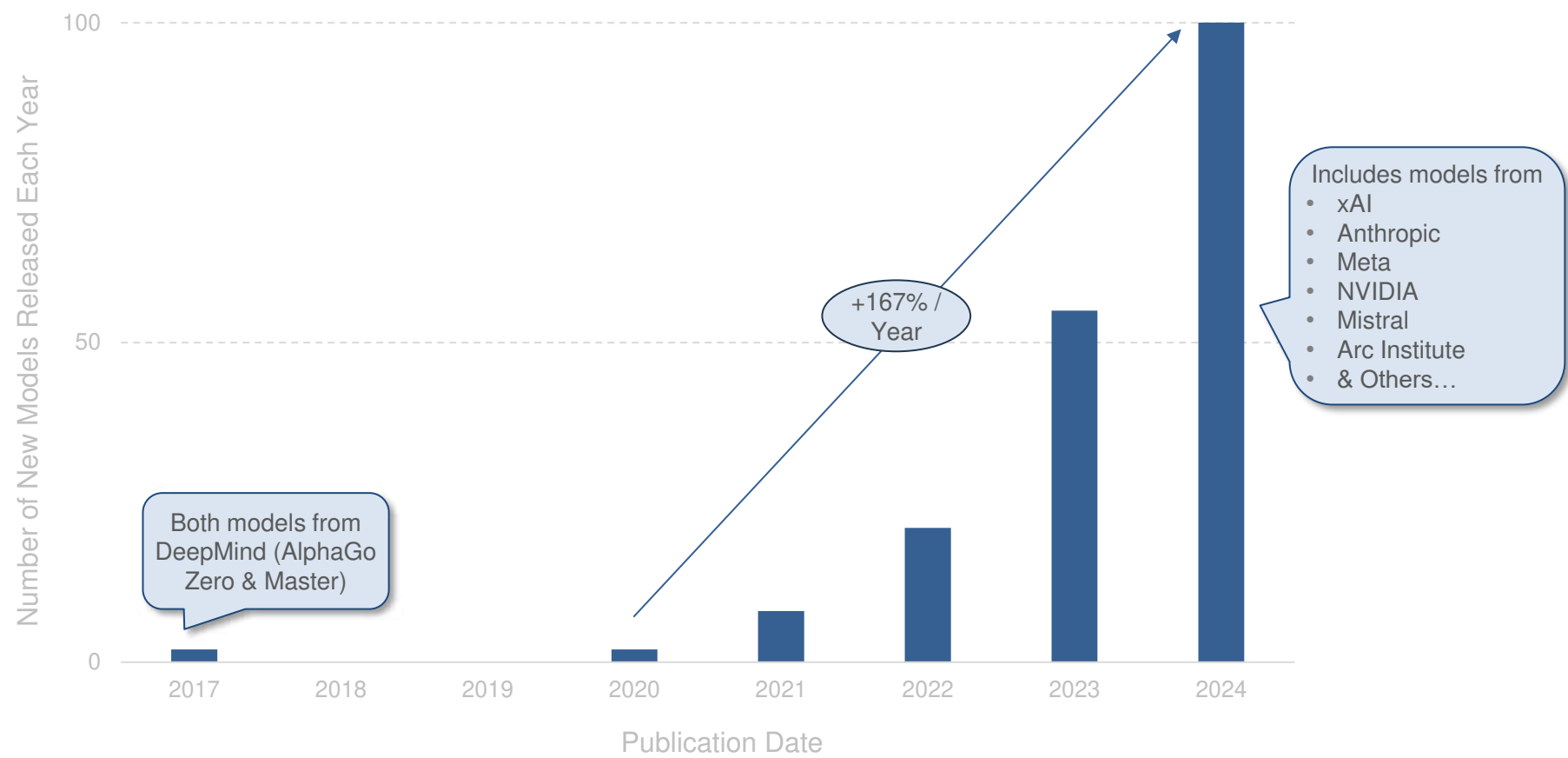
领先的 AI 超级计算机的性能 (FLOP/s) – 2019-2025, 来源: Epoch AI



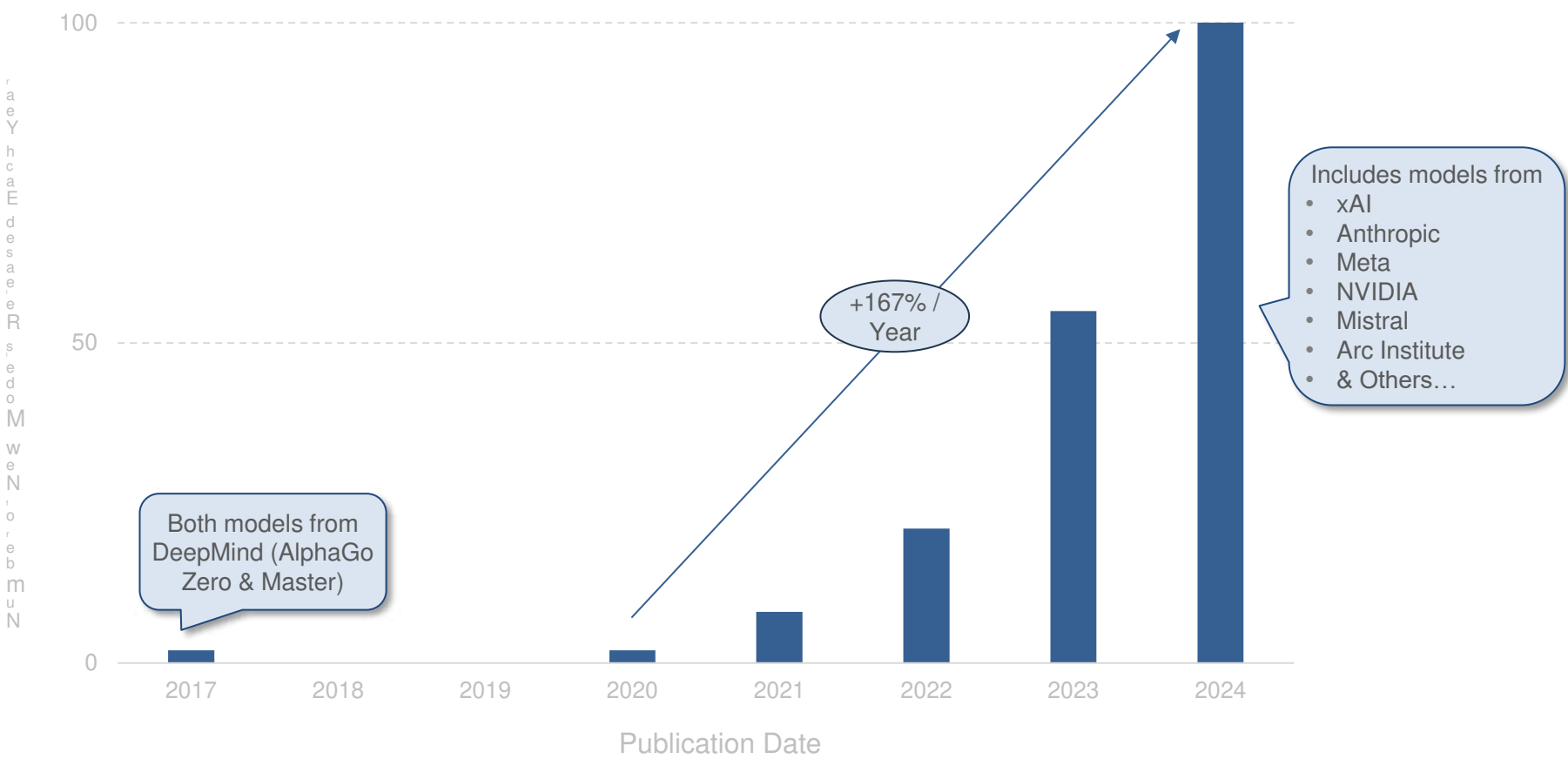
...167% Annual Growth Over Four Years in...  
Number of Powerful AI Models

… 四年内年增长率达 167%… 强大 AI 模型数量

Number of New Large-Scale AI Models (Larger than 10<sup>23</sup> FLOP\*) – 2017-2024, per Epoch AI



Number of New Large-Scale AI Models (Larger than 10<sup>23</sup> FLOP\*) – 2017-2024, per Epoch AI

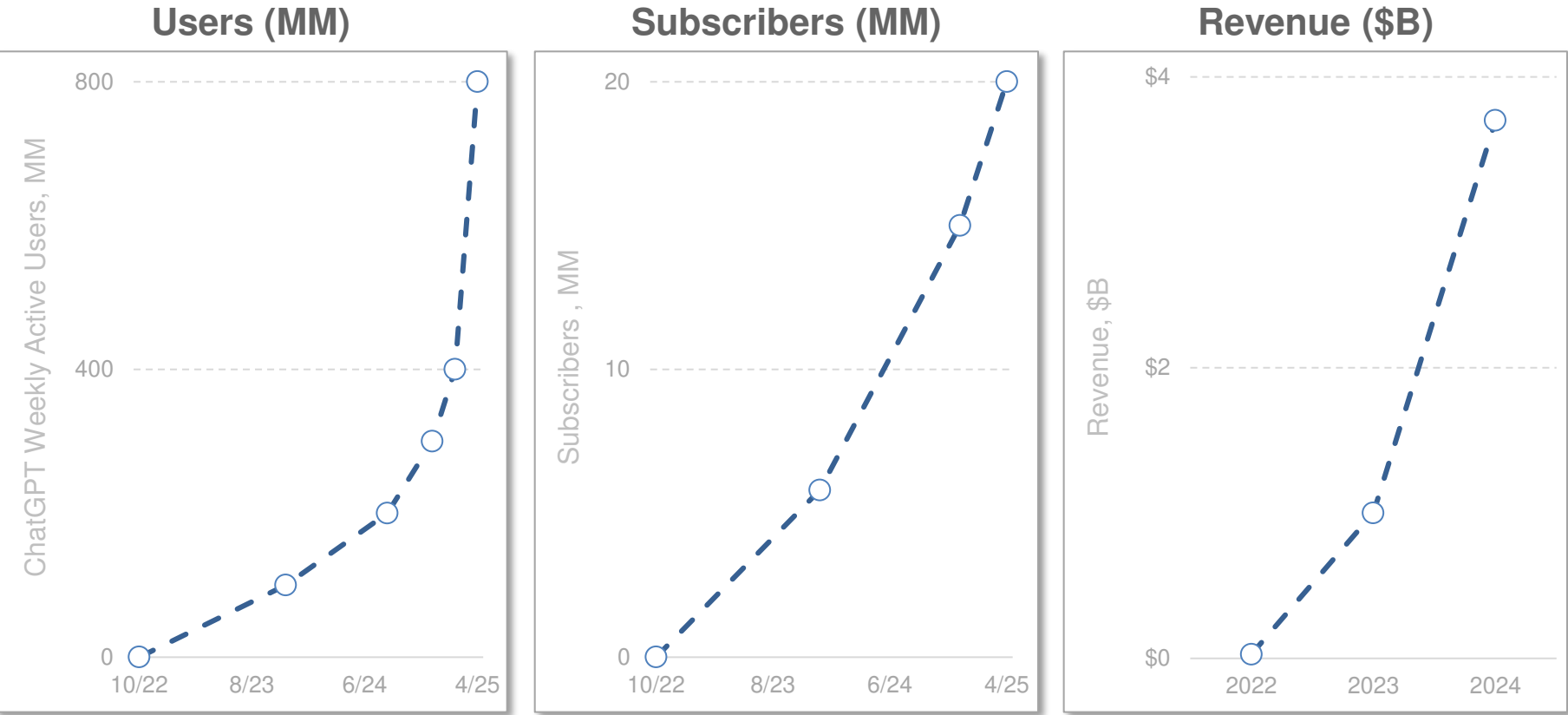


\*As of 4/25, 'Large-Scale AI Models' are generally defined as those with a training compute of 10<sup>23</sup> FLOPs or greater, per Epoch AI.  
Source: Epoch AI (5/25)

\* 截至 4 月 25 日, “大型 - 规模 AI 模型” 通常定义为每次 Epoch AI 的训练计算量达到 10<sup>23</sup> FLOP 或更高。资料来源: Epoch AI (5/25)

# ChatGPT AI User + Subscriber + Revenue Growth Ramps = Hard to Match, Ever

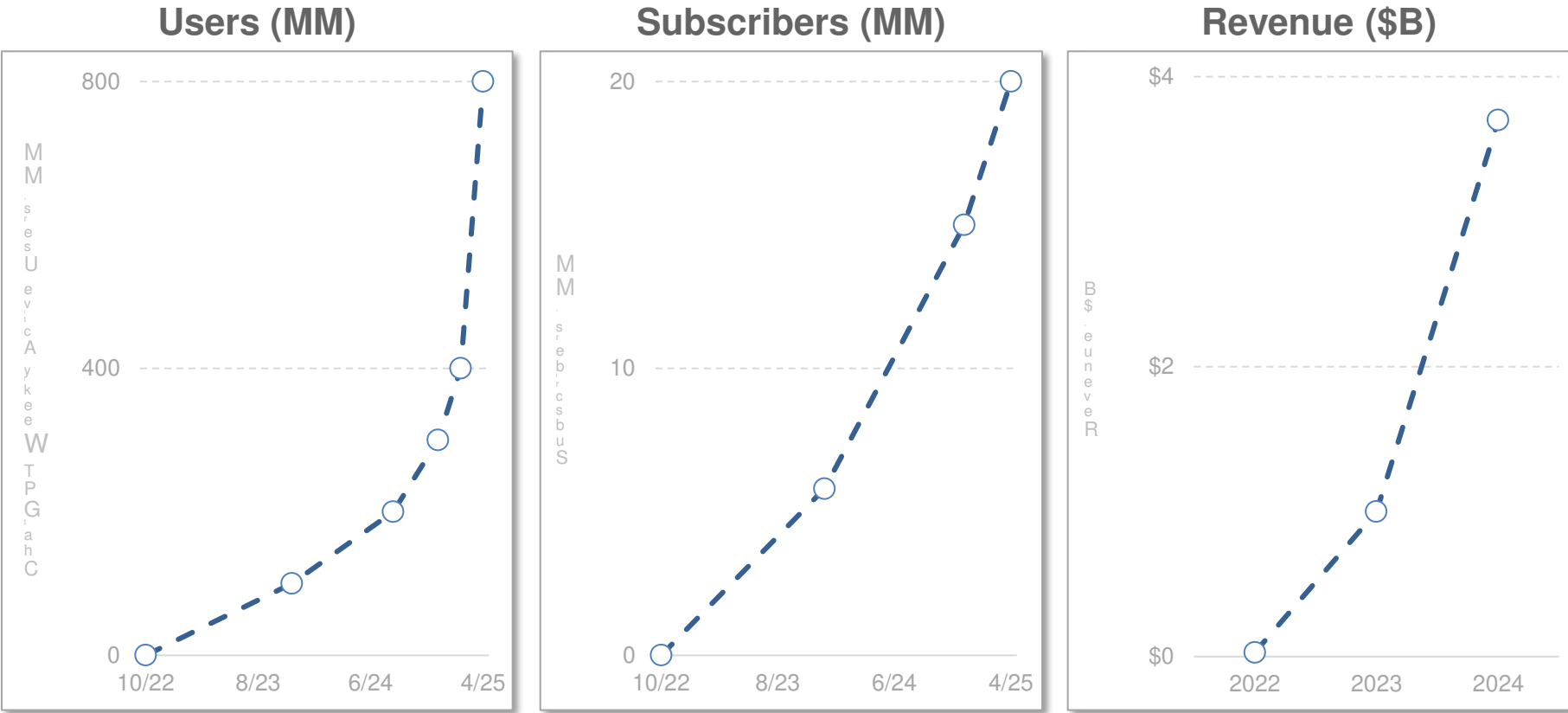
ChatGPT User + Subscriber + Revenue Growth – 10/22-4/25, per OpenAI & *The Information*



Note: 4/25 user count estimate from OpenAI CEO Sam Altman's 4/11/25 TED Talk disclosure. Revenue figures are estimates based off OpenAI disclosures. Source: OpenAI disclosures (as of 4/25), The Information (4/25) ([link](#), [link](#), [link](#) & [link](#))

# ChatGPT AI User + Subscriber + Revenue Growth Ramps = Hard to Match, Ever

ChatGPT 用户 + 订阅者 + 收入增长 – 10/22-4/25, 根据 OpenAI 和 The Information



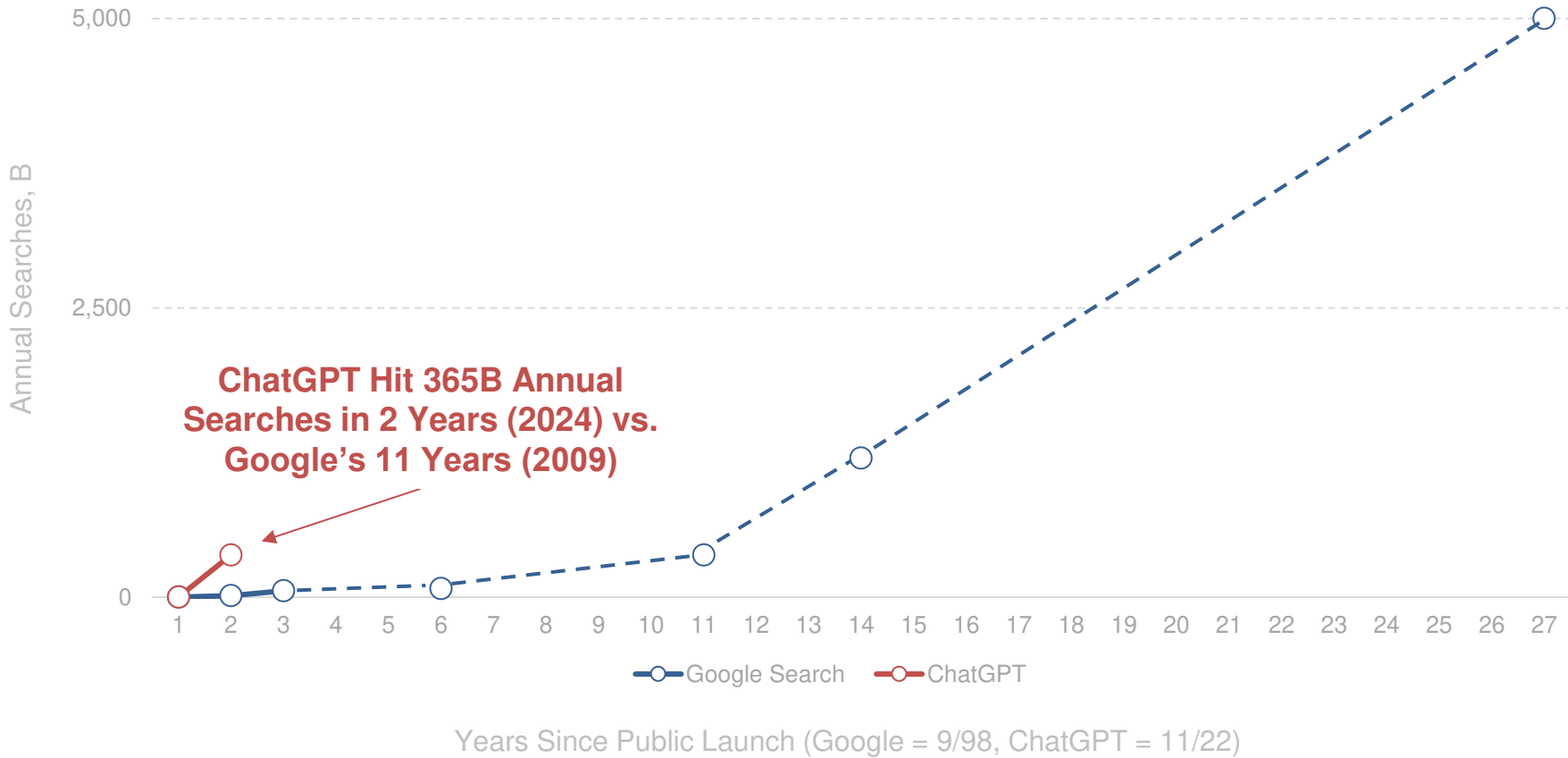
注意：4/25 的用户数量估计来自 OpenAI 首席执行官 Sam Altman 在 4/11/25 TED 演讲中的披露。收入数字是基于 OpenAI 披露的估计。来源：OpenAI 披露（截至 4/25），The Information（4/25）（[链接](#)，[链接](#)，[链接](#)和[链接](#)）



Time to 365B Annual Searches =  
ChatGPT 5.5x Faster vs. Google

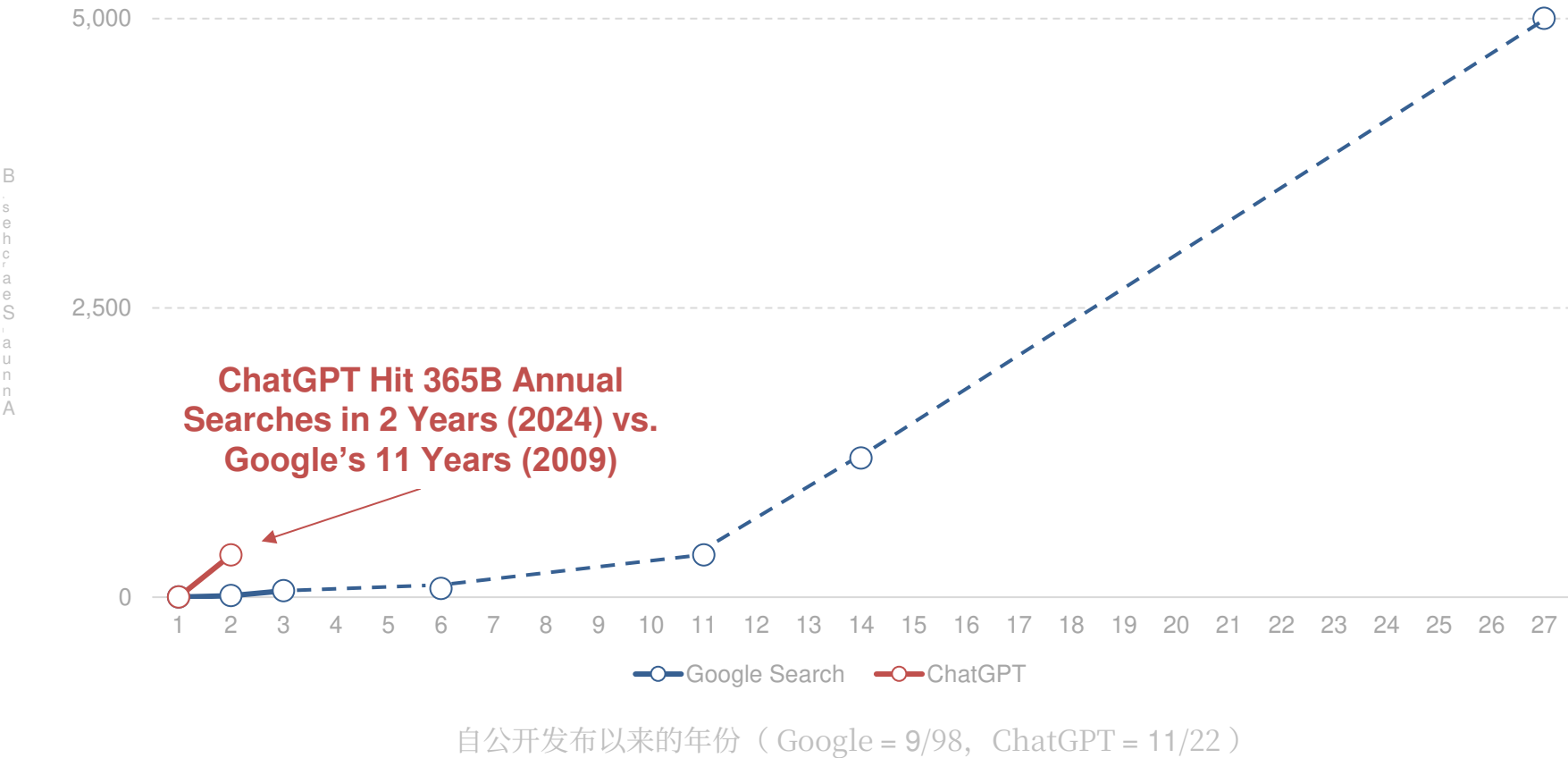
达到 365B 年度搜索量的时间 =  
ChatGPT 比 Google 快 5.5 倍

Annual Searches by Year (B) Since Public Launches of Google & ChatGPT – 1998-2025,  
per Google & OpenAI



Note: Dashed-line bars are for years where Google did not disclose annual search volumes. Source: Google public disclosures, OpenAI (12/24). ChatGPT figures are estimates per company disclosures of ~1B daily queries

自 Google 和 ChatGPT 公开发布以来的年度搜索量（ B ） – 1998-2025，数据来源：Google 和 OpenAI



注：虚线柱表示 Google 未披露年度搜索量的年份。来源：Google 公开披露，OpenAI (12/24)。ChatGPT 的数据是根据公司披露的 ~1B 每日查询量估算的

In 1998, tapping emerging Internet access, Google set out to  
‘organize the world’s information and make it  
universally accessible and useful.’

Nearly three decades later  
– after some of the fastest change humankind has seen –  
a lot of information is indeed digitized / accessible / useful.

The AI-driven evolution of how we  
access and move information is happening much faster...

...AI is a compounder – on internet infrastructure, which allows  
for wicked-fast adoption of easy-to-use broad-interest services.

1998 年，凭借新兴的互联网接入，Google 开始 “整理世界  
上的信息，使其普遍可访问且有用”。

近三十年后 – 在人类所见证的一些最快速的变化之后 –  
大量信息确实已数字化 / 可访问 / 有用。

人工智能驱动的我们移动信息方式的演  
access and 变正在发生 ch faster...

…… 人工智能是一种复合剂 —— 在互联网基础设施上，这使得易  
于使用的广泛兴趣服务能够以极快的速度被采用。

Knowledge Distribution Evolution =  
Over ~Six Centuries

知识分布演变 =  
超过 ~ 六个世纪

Knowledge Distribution – 1440-1992 =  
Static + Physical Delivery...

Printing Press – Invented 1440



Source: Wikimedia Commons

Knowledge Distribution – 1440-1992 =  
Static + Physical Delivery...

印刷机 – 发明于 1440 年



来源：维基共享资源

Internet – Public Release 1993\*



*\*The internet is widely agreed to have been ‘publicly released’ in 1993 with release of the World Wide Web (WWW) into the public domain, which allowed users to create websites; however, Tim Berners-Lee invented the World Wide Web in 1989, per CERN. Source: Google, USA Department of Defense, CERN*

互联网 – 1993 年公共发布 \*

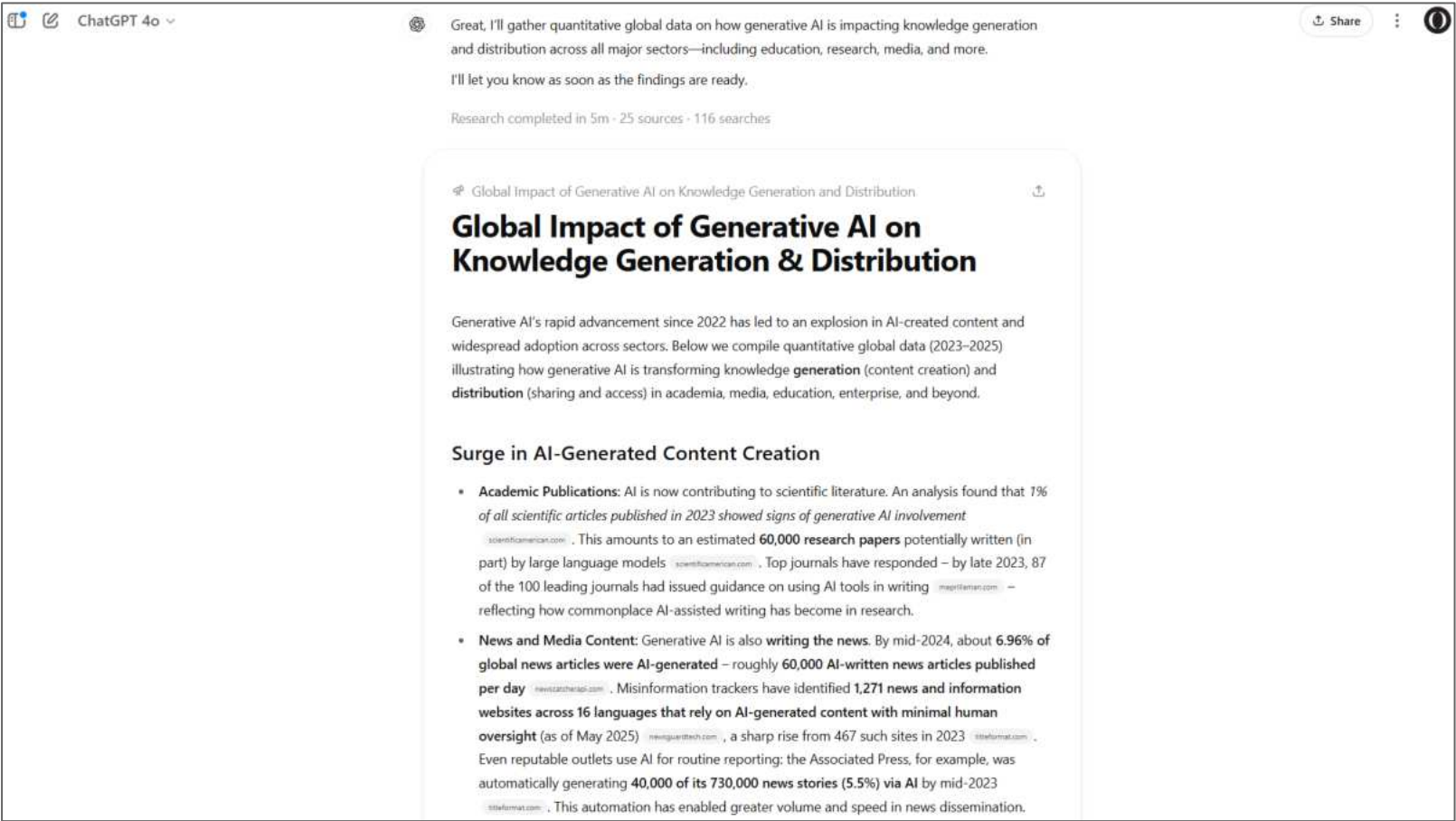


*\* 人们普遍认为，互联网于 1993 年随着万维网 (WWW) 发布到公共领域而“公开发布”，这使得用户可以创建网站；然而，根据 CERN 的说法，蒂姆·伯纳斯-李于 1989 年发明了万维网。来源：谷歌、美国国防部、欧洲核子研究中心 lic domain, which allowed users to create websites; however, Tim Berners-Lee invented the World Wide Web in 1989, per CERN. Source: Google, USA Department of Defense, CERN*

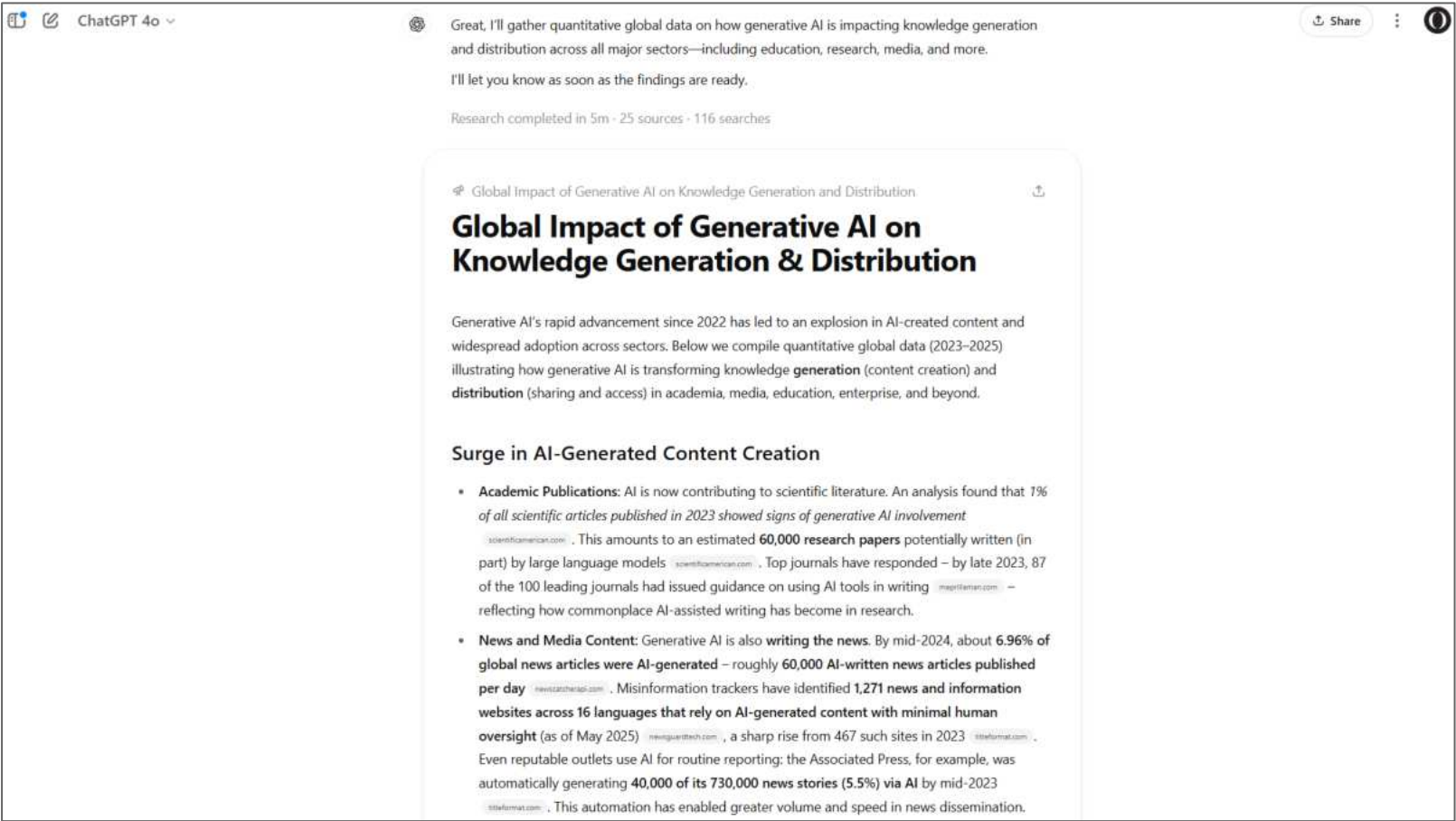
...Knowledge Distribution – 2022+ =  
Active + Digital + Generative Delivery

… 知识分发 – 2022+ = 主动 + 数字 + 生  
成交付

Generative AI – Public Launch of ChatGPT 2022\*



生成式 AI – ChatGPT 2022 年公开发布 \*



\*We define the public launch of ChatGPT in November 2022 as the public release of Generative AI which we see as AI’s ‘iPhone Moment.’ ChatGPT saw the fastest user ramp ever for a standalone product (5 days to secure 1MM users). Generative AI = AI that can create content – text, images, audio, or code – based on learned patterns. Source: OpenAI

\* 我们将 2022 年 11 月 ChatGPT 的公开发布定义为生成式 AI 的公开发布，我们将其视为 AI 的“iPhoneMoment”。ChatGPT 获得了有史以来独立产品最快的用户增长速度（5 天内获得 100 万用户）。生成式 AI = 可以创建内容的人工智能 – 文本、图像、音频或代码 – 基于学习的模式。来源：OpenAI

*Knowledge is a process of piling up facts;  
wisdom lies in their simplification.*

Martin H. Fischer, German-born American Physician / Teacher / Author (1879-1962)

知识是一个积累事实的过程；智慧在于简化它们。

Martin H. Fischer, 德裔美国医生 / 教师 / 作家（1879-1962）

AI =

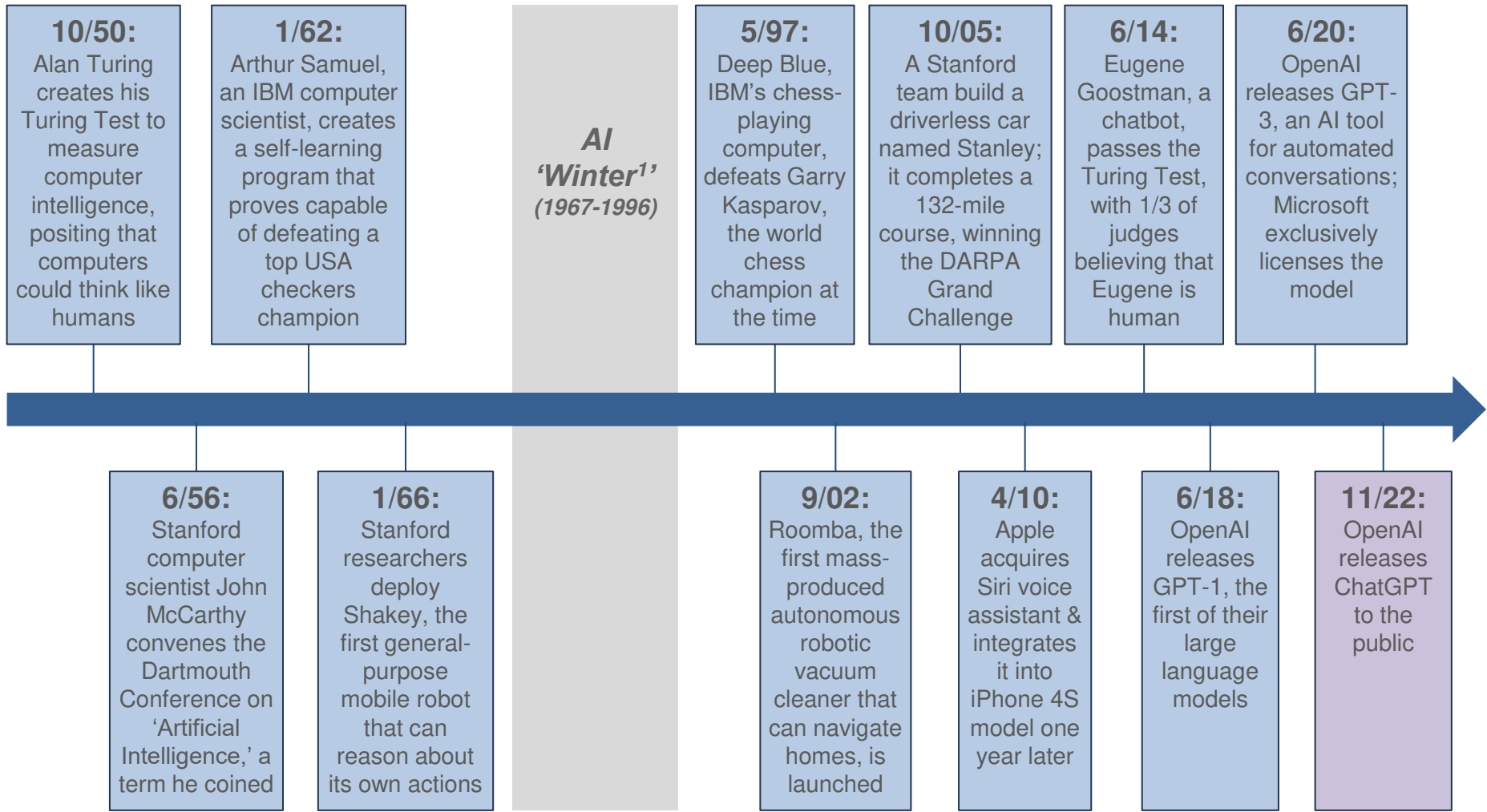
Many Years Before Lift-Off

AI =

起飞前多年

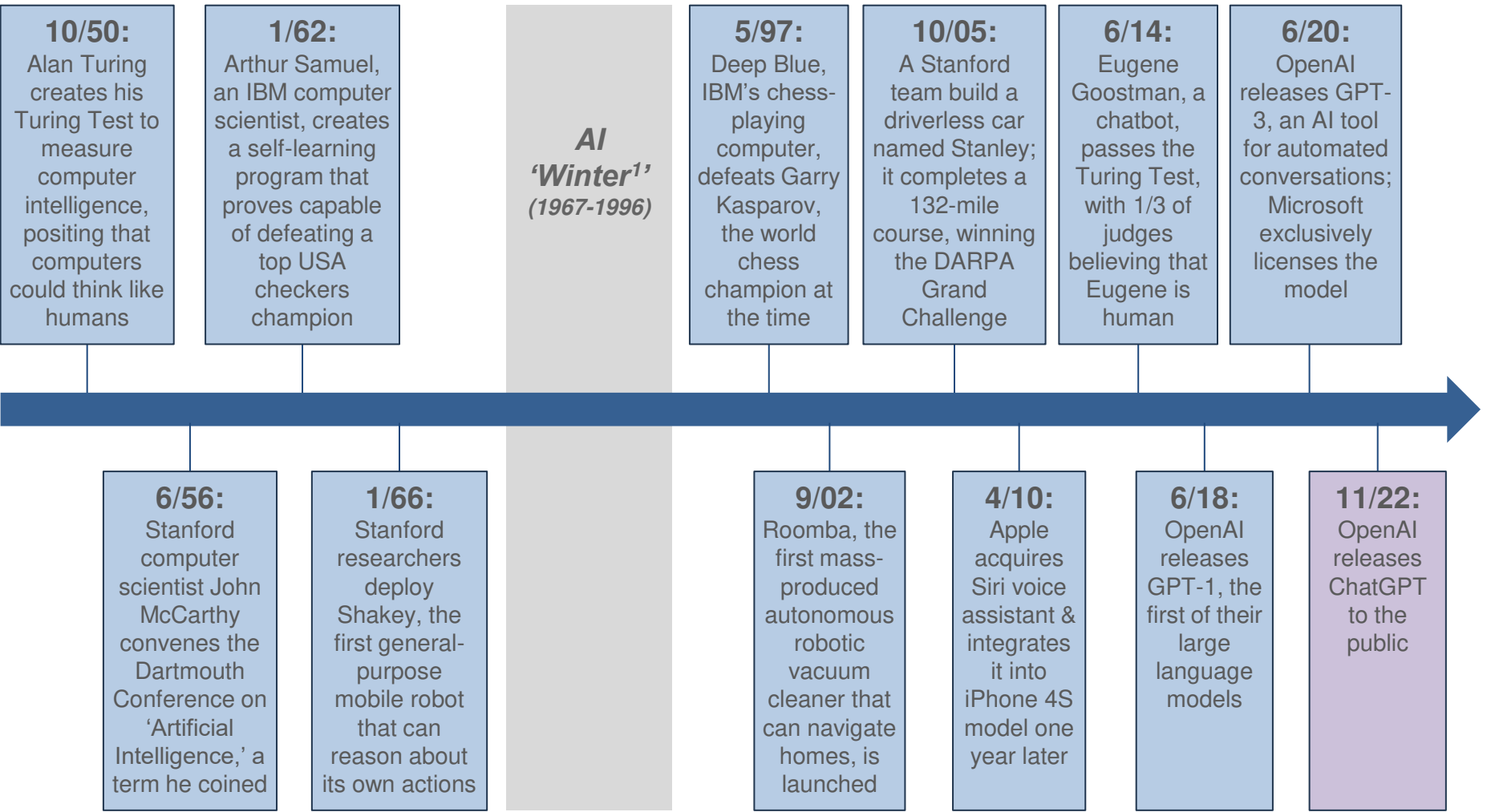


AI Milestone Timeline – 1950-2022, per Stanford University...



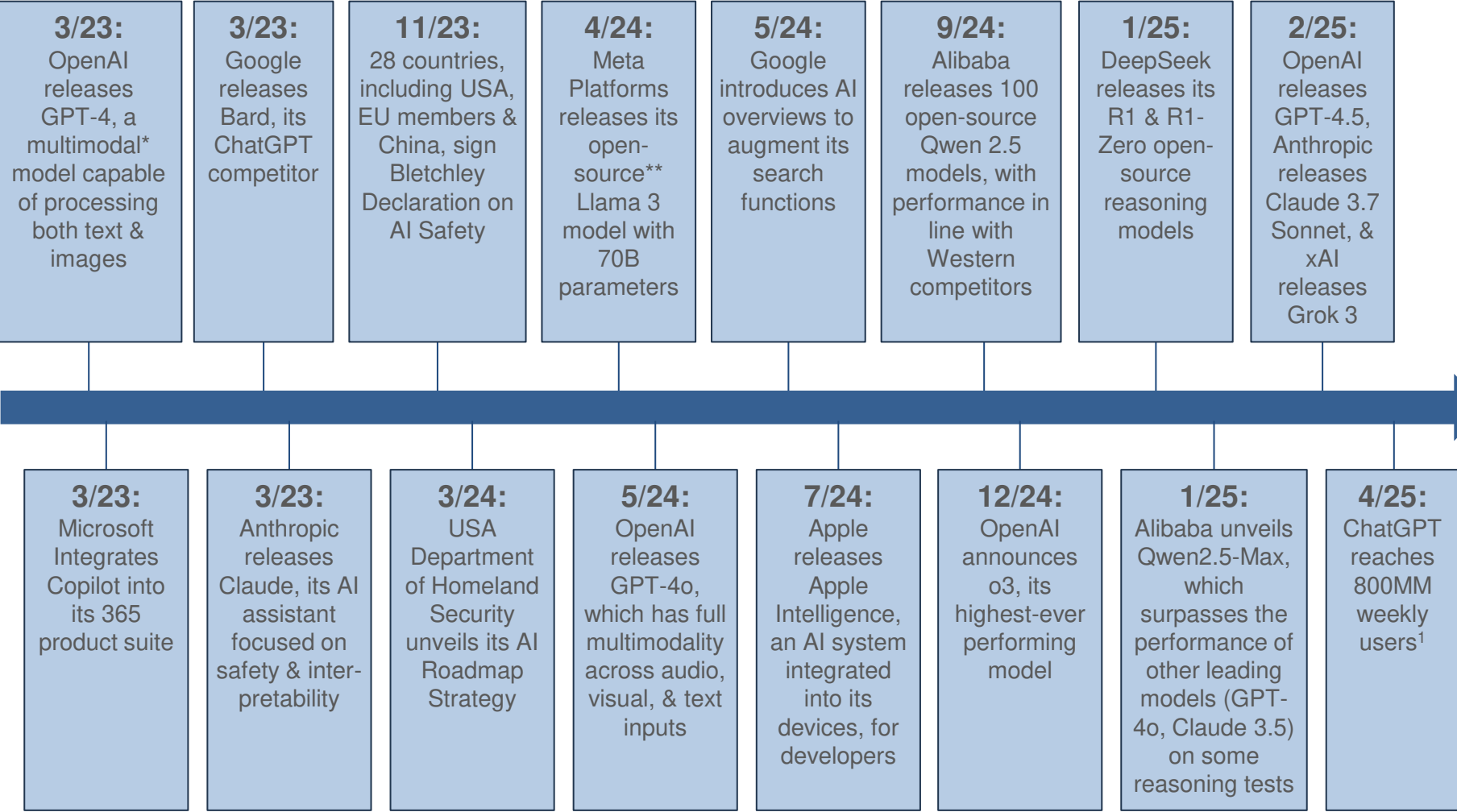
1: AI 'Winter' was a term used by Nils J. Nilsson, the Kumagai Professor of Engineering in computer science at Stanford University, to describe the period during which AI continued to make conceptual progress but could boast no significant practical successes. This subsequently led to a drop in AI interest and funding. Includes data from sources beyond Stanford. Source: Stanford University & Stanford Law School sources, iRobot, TechCrunch, BBC, OpenAI. Data aggregated by BOND.

AI Milestone Timeline – 1950-2022, 根据斯坦福大学 .....



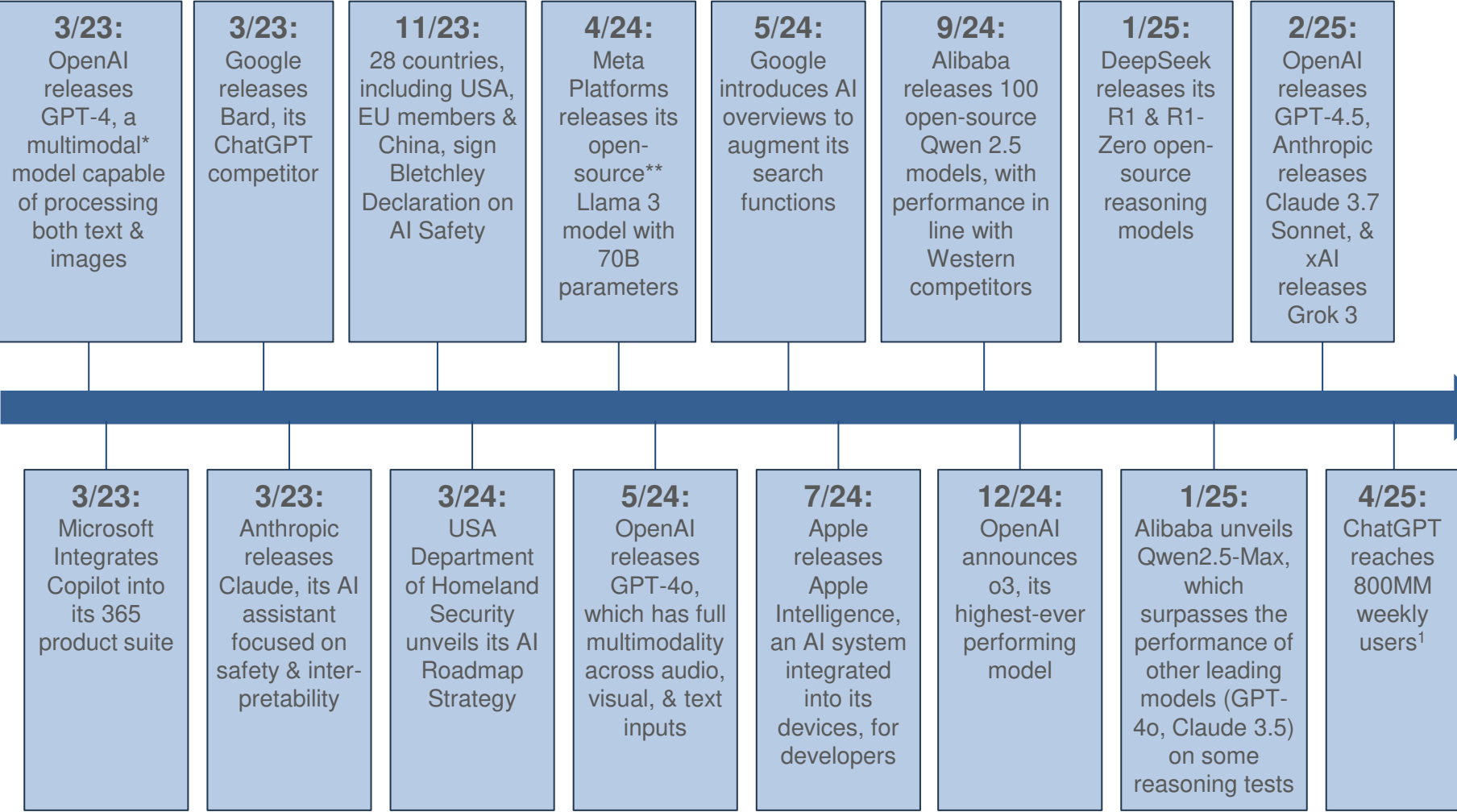
1: AI“ 寒冬 ”是 Nils J. Nilsson 使用的一个术语，他是斯坦福大学计算机科学工程学 Kumagai 教授，用来描述 AI 在这段时间内不断取得概念性进展，但未能取得任何重大实际成功的时期。这随后导致了人们对 AI 的兴趣和资金投入下降。包括来自斯坦福大学以外来源的数据。来源：斯坦福大学和斯坦福法学院来源、iRobot、TechCrunch、BBC、OpenAI。数据由 BOND 汇总。

...AI Milestone Timeline – 2023-2025, per Stanford University



\*Multimodal = AI that can understand and process multiple data types (e.g., text, images, audio) together.  
\*\*Open-source = AI models and tools made publicly available for use, modification, and redistribution.  
1) 4/25 estimate from OpenAI CEO Sam Altman's 4/11/25 TED Talk disclosure.  
Source: Aggregated by BOND from OpenAI, Microsoft, Google, Anthropic, Meta, Apple, Alibaba, Deepseek, UK Government, US Department of Homeland Security. China data may be subject to informational limitations due to government restrictions.

...AI Milestone Timeline – 2023-2025, 根据斯坦福大学



\*Multimodal = AI that can understand and process multiple data types (e.g., text, images, audio) together.  
\*\*Open-source = AI models and tools made publicly available for use, modification, and redistribution.  
1) 4/25 estimate from OpenAI CEO Sam Altman's 4/11/25 TED Talk disclosure.  
Source: Aggregated by BOND from OpenAI, Microsoft, Google, Anthropic, Meta, Apple, Alibaba, Deepseek, UK Government, US Department of Homeland Security. China data may be subject to informational limitations due to government restrictions.

AI =

Circa Q2:25

AI =

大约 Q2:25

ChatGPT 4o ▾

1. Write or edit anything

➤ Emails, essays, contracts, poems, code — instantly and fluently.

2. Summarize and explain complex material

➤ Simplifies PDFs, legal docs, research, or code into plain English.

3. Tutor you on nearly any subject

➤ Learn math, history, languages, or test prep, step by step.

4. Be your thinking partner

➤ Brainstorm ideas, debug logic, or pressure-test assumptions.

5. Automate repetitive work

➤ Generate reports, clean data, outline slides, rewrite text.

6. Roleplay anyone you need

➤ Prep for interviews, simulate customers, rehearse conversations.

7. Connect you to tools

➤ Write code for APIs, spreadsheets, calendars, or the web.

8. Offer therapy and companionship

➤ Talk through your day, reframe thoughts, or just listen.

9. Help you find purpose

➤ Clarify values, define goals, and map actions that matter.

10. Organize your life

➤ Plan trips, build routines, structure your week or workflow.

ChatGPT 4o ▾

1. Write or edit anything

➤ Emails, essays, contracts, poems, code — instantly and fluently.

2. Summarize and explain complex material

➤ Simplifies PDFs, legal docs, research, or code into plain English.

3. Tutor you on nearly any subject

➤ Learn math, history, languages, or test prep, step by step.

4. Be your thinking partner

➤ Brainstorm ideas, debug logic, or pressure-test assumptions.

5. Automate repetitive work

➤ Generate reports, clean data, outline slides, rewrite text.

6. Roleplay anyone you need

➤ Prep for interviews, simulate customers, rehearse conversations.

7. Connect you to tools

➤ Write code for APIs, spreadsheets, calendars, or the web.

8. Offer therapy and companionship

➤ Talk through your day, reframe thoughts, or just listen.

9. Help you find purpose

➤ Clarify values, define goals, and map actions that matter.

10. Organize your life

➤ Plan trips, build routines, structure your week or workflow.

AI =

Circa 2030?

AI =

大约 2030 年?

# Top Ten Things AI Will Likely Do in Five Years, per ChatGPT

ChatGPT 4o ▾

1. Generate human-level text, code & logic

► Chatbots, software engineering, business plans, legal analysis.

2. Create full-length films & games

► Scripts, characters, scenes, gameplay mechanics, voice acting.

3. Understand & speak like a human

► Emotionally aware assistants, real-time multilingual voice agents.

4. Power advanced personal assistants

► Life planning, memory recall, coordination across all apps and devices.

5. Operate humanlike robots

► Household helpers, elderly care, retail and hospitality automation.

6. Run autonomous customer service & sales

► End-to-end resolution, upselling, CRM integrations, 24/7 support.

7. Personalize entire digital lives

► Adaptive learning, dynamic content curation, individualized health coaching.

8. Build and run autonomous businesses

► AI-driven startups, inventory and pricing optimization, full digital operations.

9. Drive autonomous discovery in science

► Drug design, materials synthesis, climate modeling, novel hypothesis testing.

10. Collaborate creatively like a partner

► Co-writing novels, music production, fashion design, architecture.

Source: ChatGPT (5/15/25)

# Top Ten Things AI Will Likely Do in Five Years, per ChatGPT

ChatGPT 4o ▾

1. Generate human-level text, code & logic

► Chatbots, software engineering, business plans, legal analysis.

2. Create full-length films & games

► Scripts, characters, scenes, gameplay mechanics, voice acting.

3. Understand & speak like a human

► Emotionally aware assistants, real-time multilingual voice agents.

4. Power advanced personal assistants

► Life planning, memory recall, coordination across all apps and devices.

5. Operate humanlike robots

► Household helpers, elderly care, retail and hospitality automation.

6. Run autonomous customer service & sales

► End-to-end resolution, upselling, CRM integrations, 24/7 support.

7. Personalize entire digital lives

► Adaptive learning, dynamic content curation, individualized health coaching.

8. Build and run autonomous businesses

► AI-driven startups, inventory and pricing optimization, full digital operations.

9. Drive autonomous discovery in science

► Drug design, materials synthesis, climate modeling, novel hypothesis testing.

10. Collaborate creatively like a partner

► Co-writing novels, music production, fashion design, architecture.

Source: ChatGPT (5/15/25)

AI =

Circa 2035?

AI =

大约 2035 年?

# Top Ten Things AI Will Likely Do in Ten Years, per ChatGPT

ChatGPT 4o ▾

1. Conduct scientific research

➤ Generate hypotheses, run simulations, design and analyze experiments.

2. Design advanced technologies

➤ Discover materials, engineer biotech, and prototype energy systems.

3. Simulate human-like minds

➤ Create digital personas with memory, emotion, and adaptive behavior.

4. Operate autonomous companies

➤ Manage R&D, finance, and logistics with minimal human input.

5. Perform complex physical tasks

➤ Handle tools, assemble components, and adapt in real-world spaces.

6. Coordinate systems globally

➤ Optimize logistics, energy use, and crisis response at scale.

7. Model full biological systems

➤ Simulate cells, genes, and organisms for research and therapy.

8. Offer expert-level decisions

➤ Deliver real-time legal, medical, and business advice.

9. Shape public debate and policy

➤ Moderate forums, propose laws, and balance competing interests.

10. Build immersive virtual worlds

➤ Generate interactive 3D environments directly from text prompts.

Source: ChatGPT (5/15/25)

# Top Ten Things AI Will Likely Do in Ten Years, per ChatGPT

ChatGPT 4o ▾

1. Conduct scientific research

➤ Generate hypotheses, run simulations, design and analyze experiments.

2. Design advanced technologies

➤ Discover materials, engineer biotech, and prototype energy systems.

3. Simulate human-like minds

➤ Create digital personas with memory, emotion, and adaptive behavior.

4. Operate autonomous companies

➤ Manage R&D, finance, and logistics with minimal human input.

5. Perform complex physical tasks

➤ Handle tools, assemble components, and adapt in real-world spaces.

6. Coordinate systems globally

➤ Optimize logistics, energy use, and crisis response at scale.

7. Model full biological systems

➤ Simulate cells, genes, and organisms for research and therapy.

8. Offer expert-level decisions

➤ Deliver real-time legal, medical, and business advice.

9. Shape public debate and policy

➤ Moderate forums, propose laws, and balance competing interests.

10. Build immersive virtual worlds

➤ Generate interactive 3D environments directly from text prompts.

Source: ChatGPT (5/15/25)



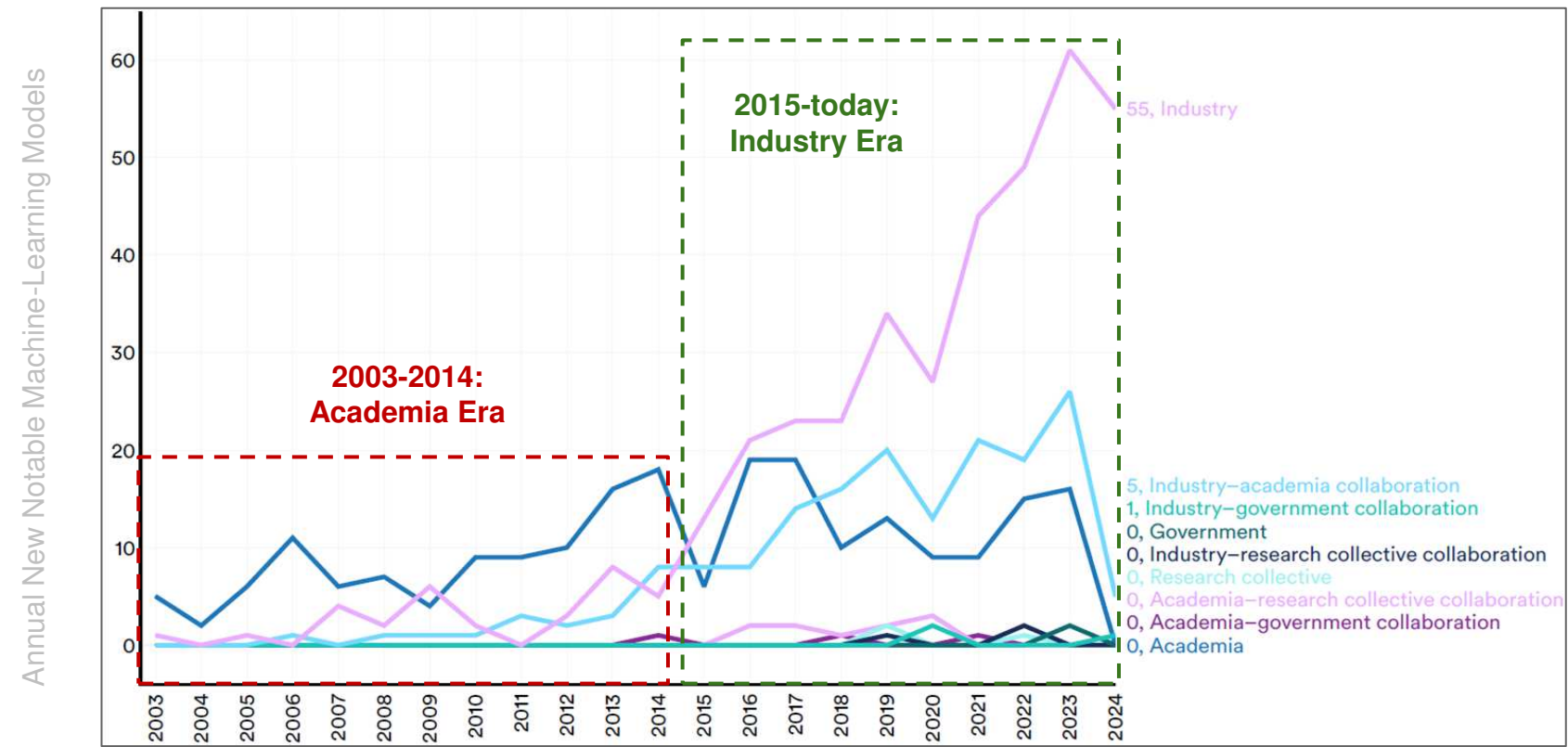
AI Development Trending =  
Unprecedented

AI 发展趋势 =  
前所未有

# Machine-Learning Model\* Trending = In 2015... Industry Surpassed Academia as Data + Compute + Financial Needs Rose

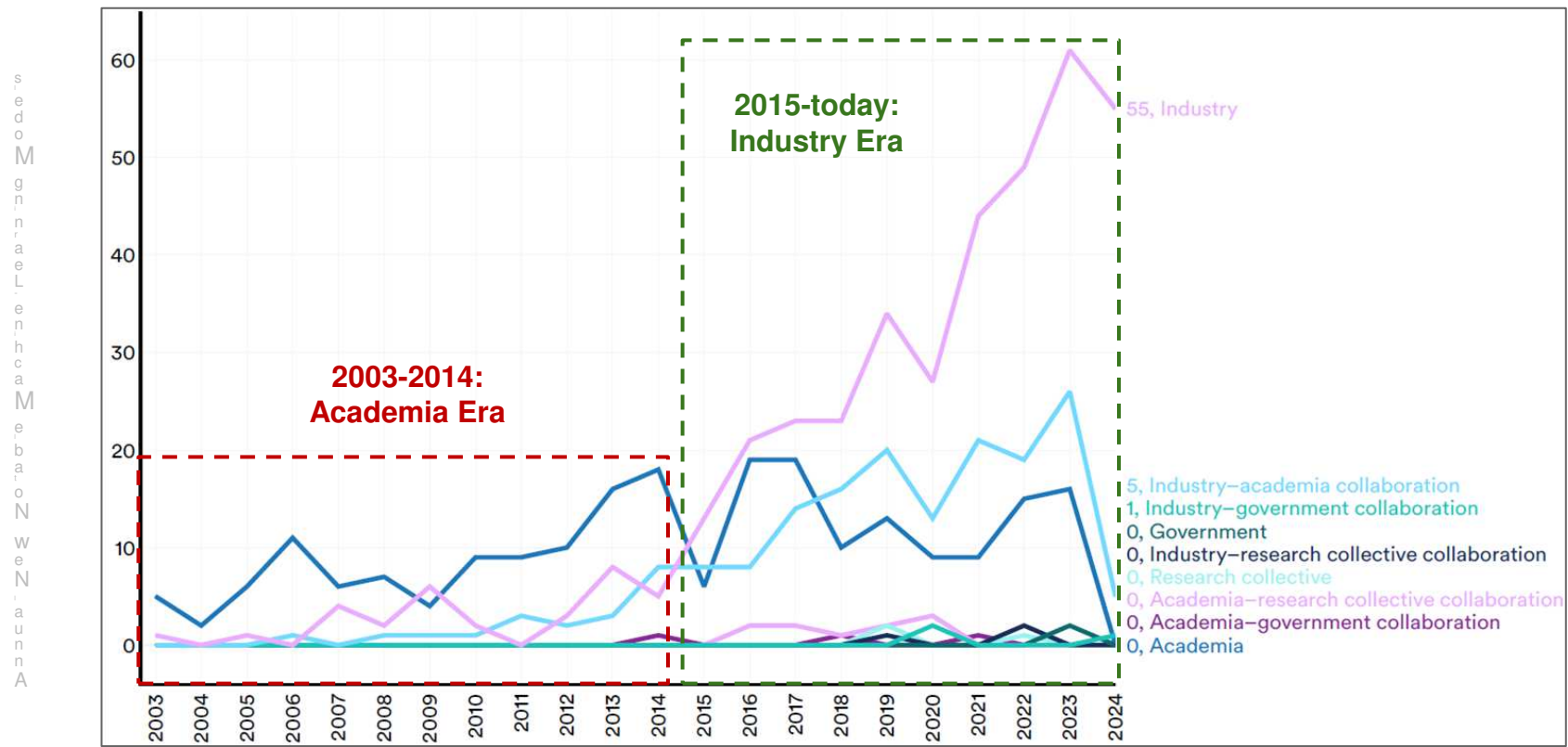
# 机器学习模型 \* 趋势 = 2015 年 ... 随着数据 + 计算 + 财务需求的增长，行业超越了学术界

Global Notable Machine Learning Models by Sector – 2003-2024, per Stanford HAI



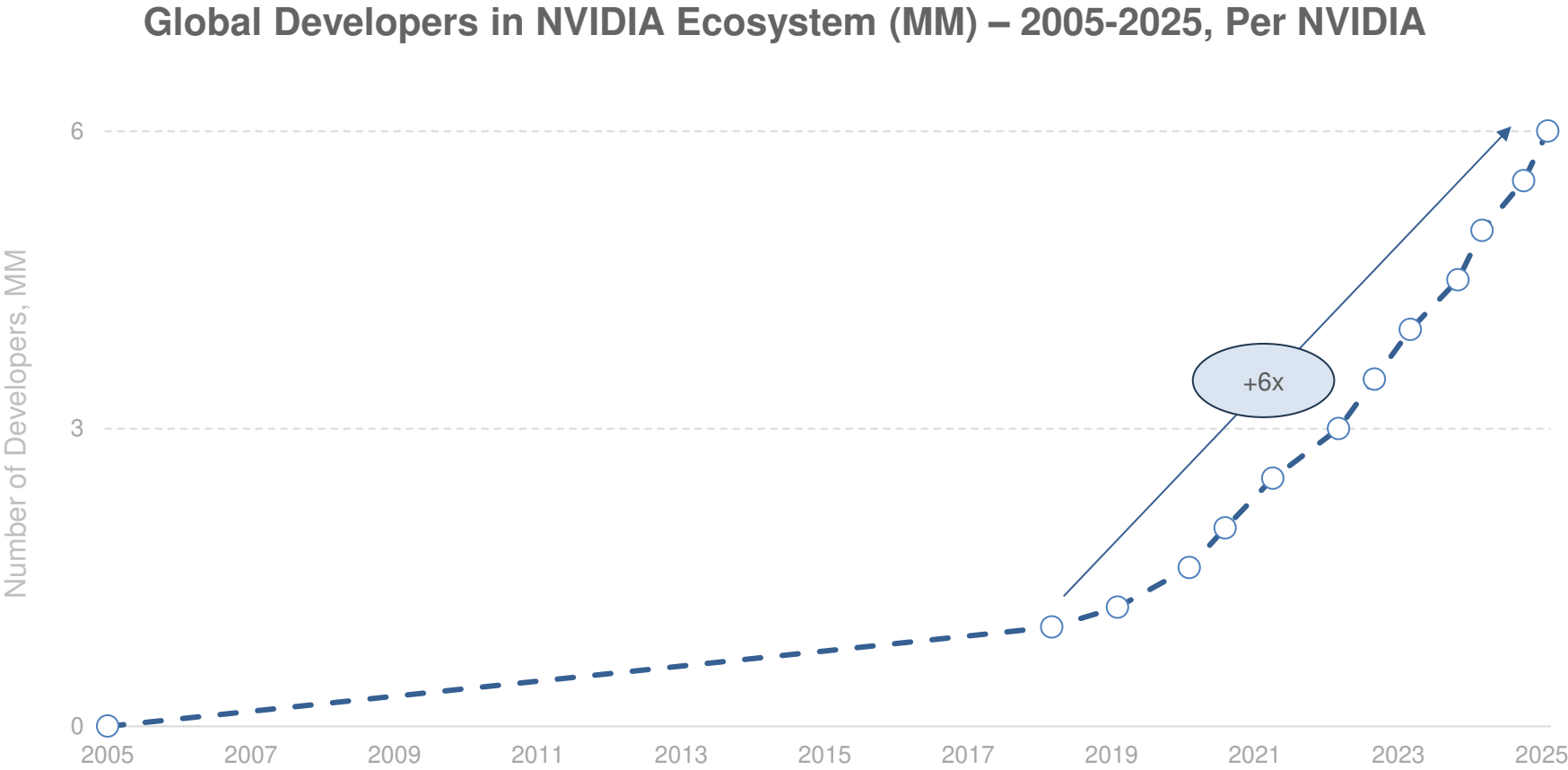
\*Machine Learning = A subset of AI where machines learn from patterns in data without being explicitly programmed.  
Note: Academia includes models developed by one or more institutions, including government agencies. Industry-academia collaboration excludes government partnerships and only captures partnerships between academic institutions and industry. Industry excludes models developed in partnership with any entity other than another company. Epoch AI, an AI Index data provider, uses the term 'notable machine learning models' to designate particularly influential models within the AI/machine learning ecosystem. Epoch maintains a database of 900 AI models released since the 1950s, selecting entries based on criteria such as state-of-the-art advancements, historical significance, or high citation rates. Since Epoch manually curates the data, some models considered notable by some may not be included. A count of zero academic models does not mean that no notable models were produced by academic institutions in 2023, but rather that Epoch AI has not identified any as notable. Additionally, academic publications often take longer to gain recognition, as highly cited papers introducing significant architectures may take years to achieve prominence. China data may be subject to informational limitations due to government restrictions. Source: Nestor Maslej et al., 'The AI Index 2025 Annual Report,' AI Index Steering Committee, Stanford HAI (4/25)

Global Notable Machine Learning Models by Sector – 2003-2024, per Stanford HAI



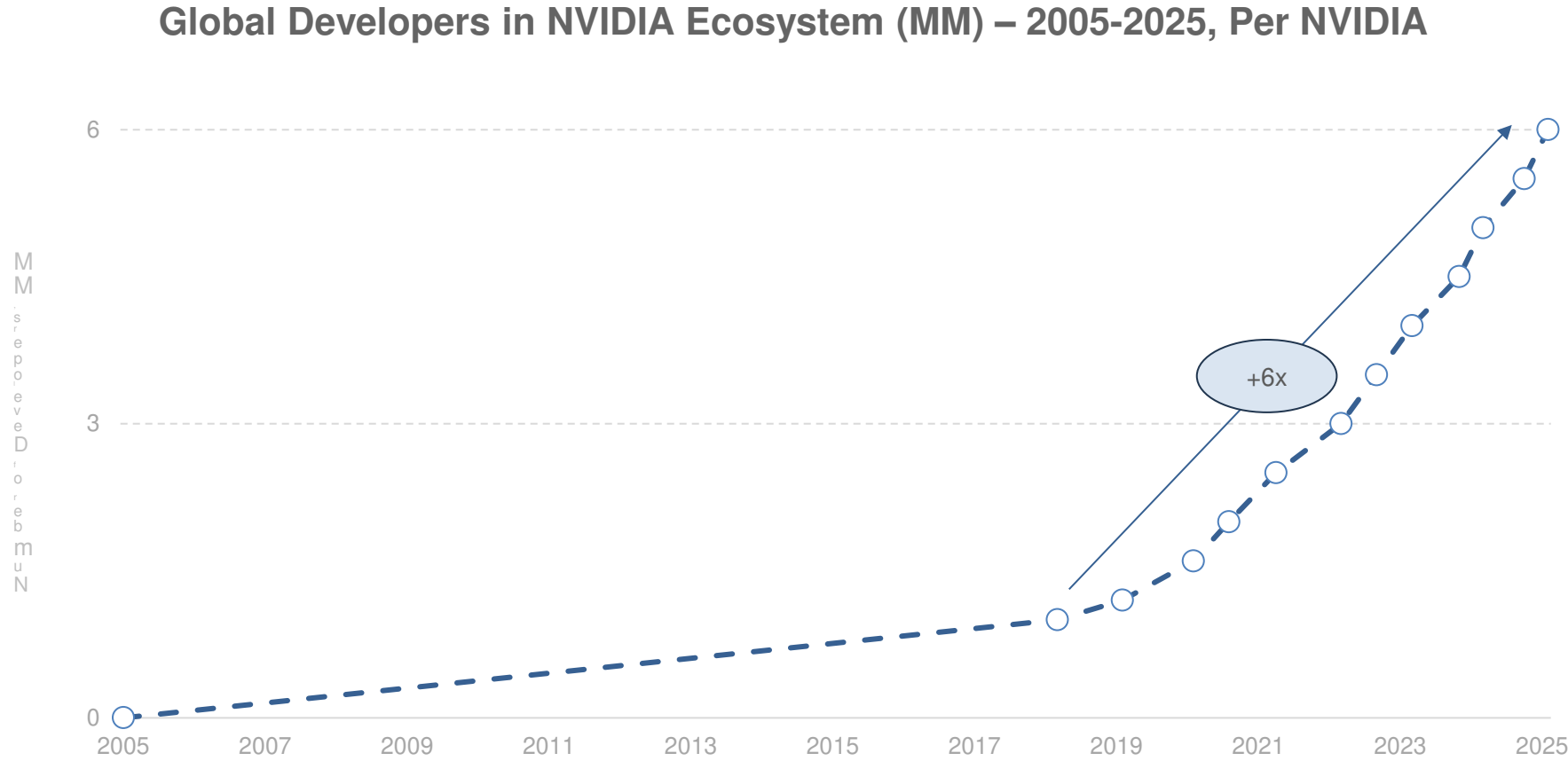
\* 机器学习 = 人工智能的一个子集，机器通过数据中的模式进行学习，而无需显式编程。注：学术界包括由一个或多个机构（包括政府机构）开发的模型。产业界 - 学术界合作不包括政府合作，仅包括学术机构与产业界之间的合作。产业界不包括与其他公司以外的任何实体合作开发的模型。Epoch AI 是一家 AI 指数数据提供商，使用“著名机器学习模型”一词来指定 AI/ 机器学习生态系统中特别有影响力的模型。Epoch 维护着一个自 20 世纪 50 年代以来发布的 900 个人工智能模型数据库，根据诸如最先进的进步、历史意义或高引用率等标准选择条目。由于 Epoch 手动管理数据，因此某些人认为值得注意的一些模型可能未包含在内。零学术模型计数并不意味着 2023 年学术机构未产生任何值得注意的模型，而是 Epoch AI 尚未将任何模型识别为值得注意的模型。此外，学术出版物通常需要更长的时间才能获得认可，因为引入重要架构的高引用论文可能需要数年才能获得突出地位。由于政府的限制，中国的数据可能会受到信息限制。来源：Nestor Maslej 等人，“AI 指数 2025 年度报告”，AI 指数指导委员会，斯坦福 HAI (4/25)

# AI Developer Growth (NVIDIA Ecosystem as Proxy) = +6x to 6MM Developers Over Seven Years



*Note: We assume negligible developers in NVIDIA's ecosystem in 2005 per this text from an 8/20 blog post titled '2 Million Registered Developers, Countless Breakthroughs': 'It took 13 years to reach 1 million registered developers, and less than two more to reach 2 million.' Source: NVIDIA blog posts, press releases, & company overviews*

# AI Developer Growth (NVIDIA Ecosystem as Proxy) = +6x to 6MM Developers Over Seven Years

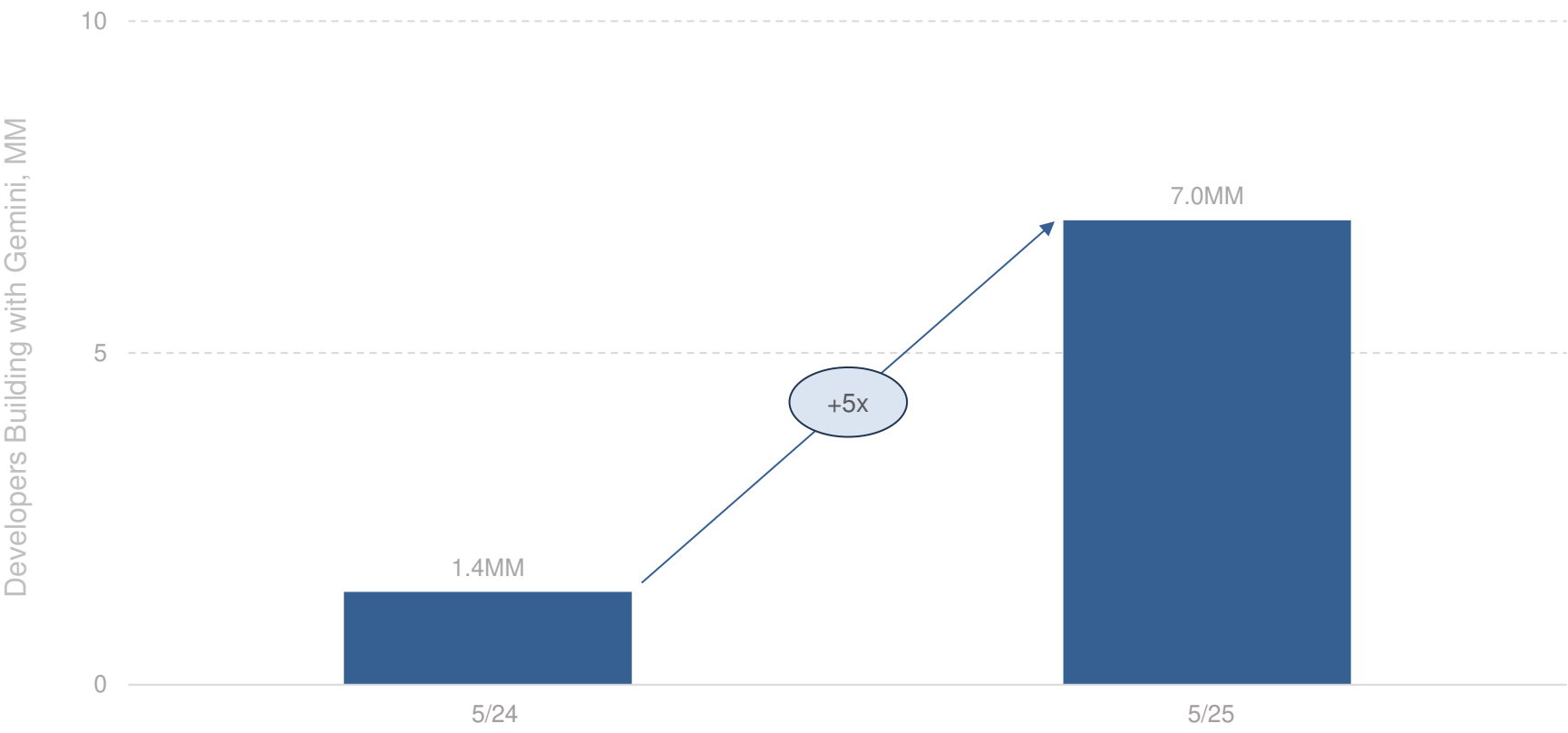


注意：根据 8 月 20 日一篇题为“2 Million Registered Developers, Countless Breakthroughs”的博文中的这段文字，我们假设 2005 年 NVIDIA 生态系统中的开发者可以忽略不计：“我们花了 13 年时间才达到 100 万注册开发者，而不到两年就达到了 200 万。”来源：NVIDIA 博客文章、新闻稿和公司概况

AI Developer Growth (Google Ecosystem as Proxy) = +5x to 7MM Developers Y/Y

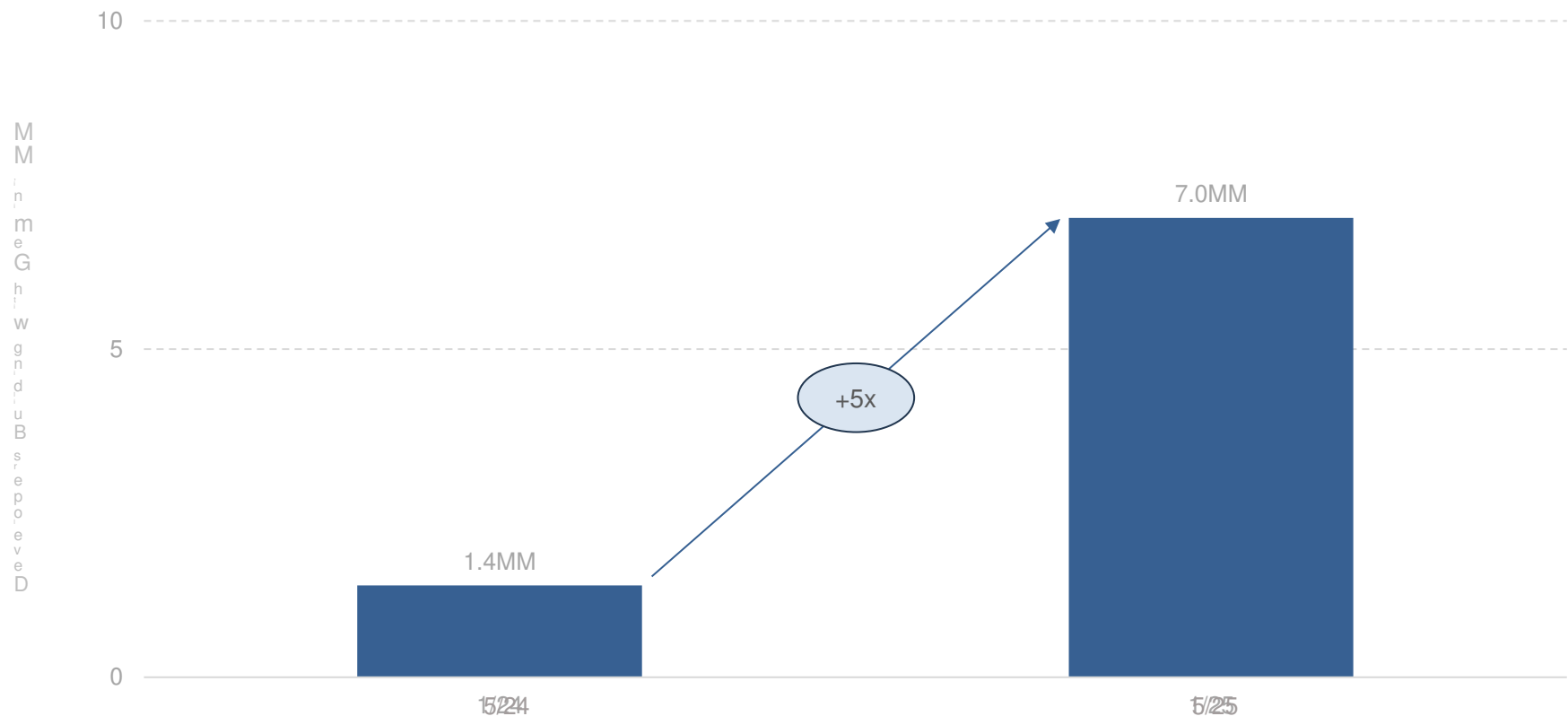
AI 开发者增长（以 Google 生态系统为代表） =+5x 到 7MM 开发者 Y/Y

Estimated Global Developers in Google Ecosystem (MM) – 5/24-5/25, Per Google



Note: Per Google in 5/25, 'Over 7 million developers are building with Gemini, five times more than this time last year.' Source: Google, 'Google I/O 2025: From research to reality' (5/25)

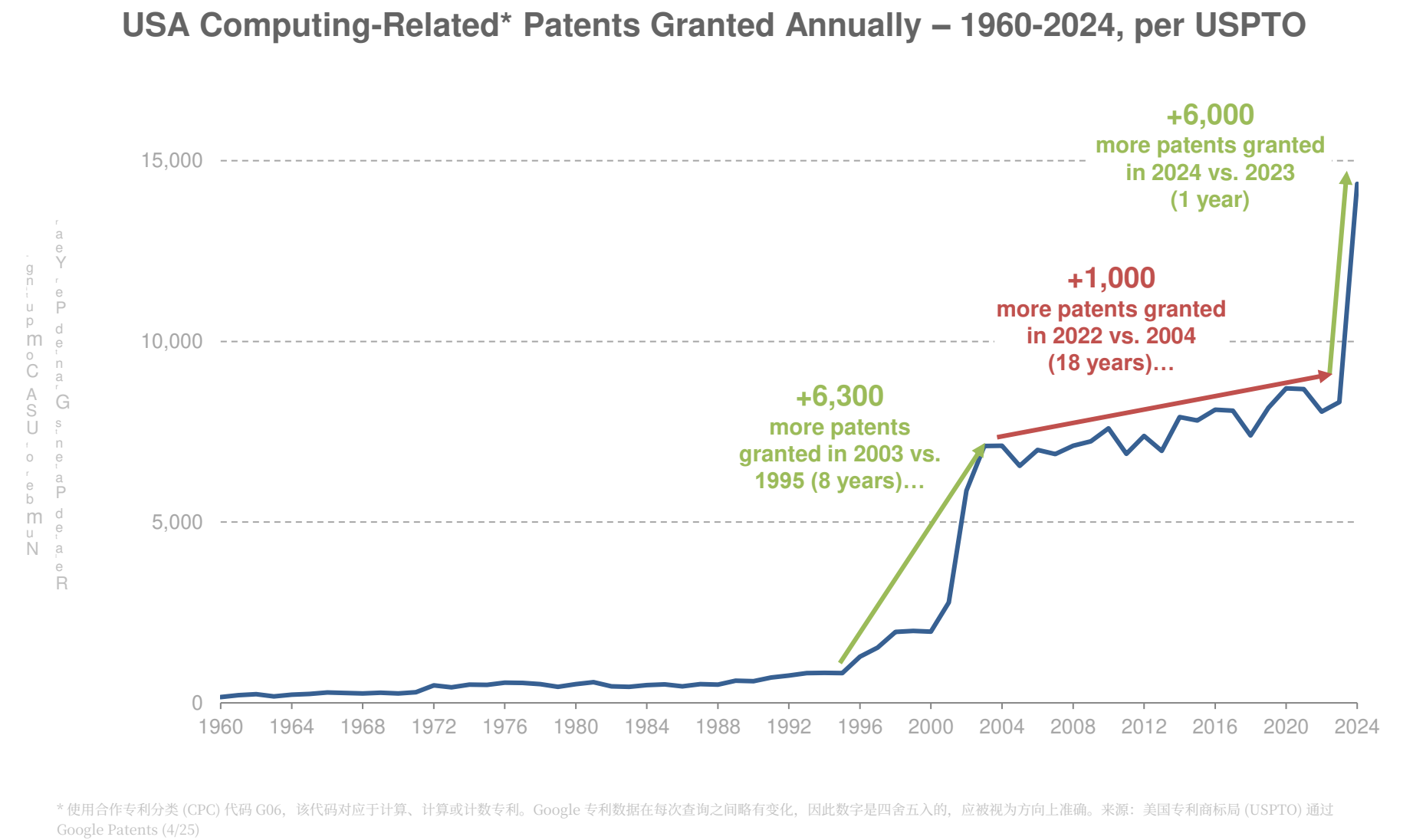
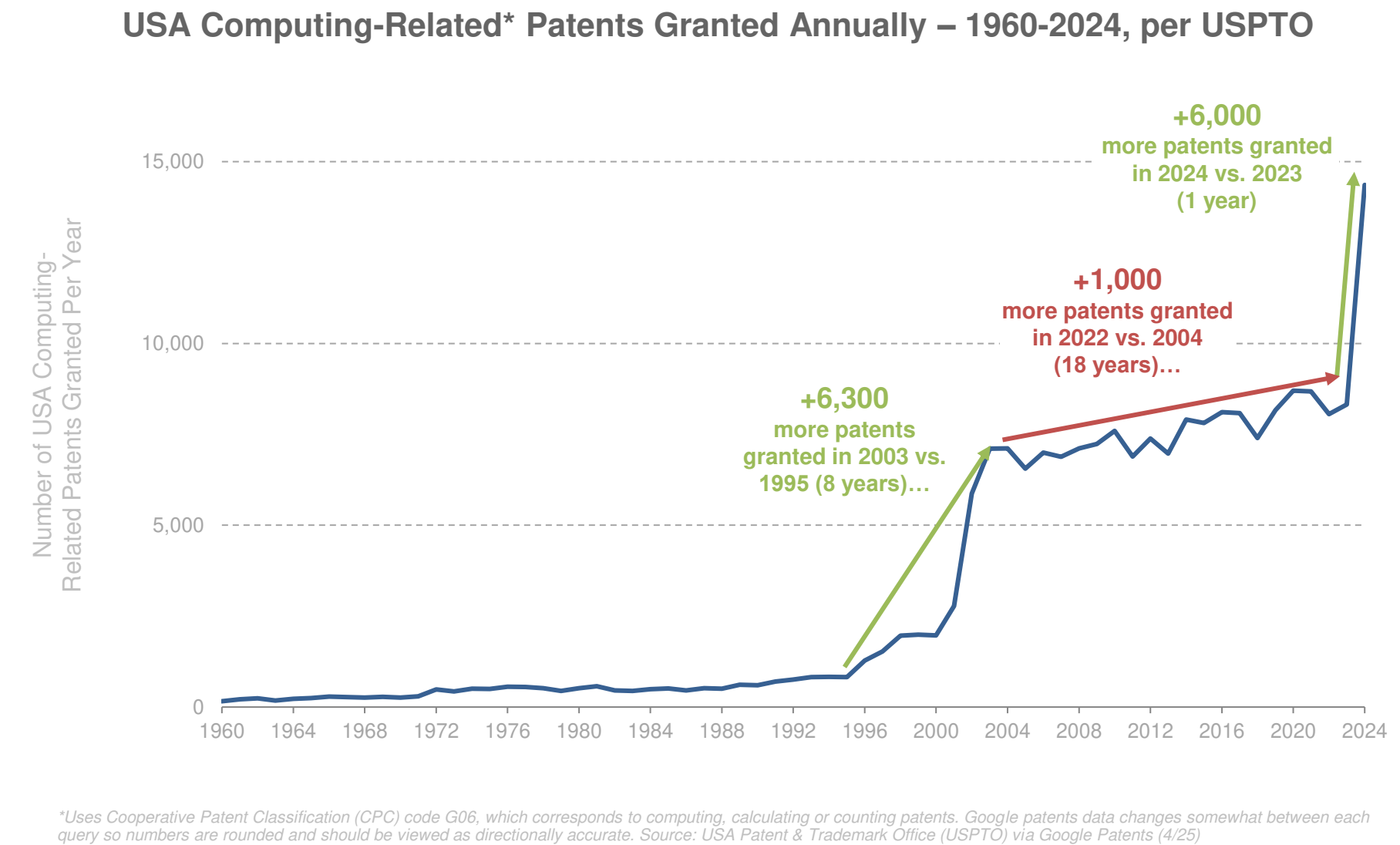
Estimated Global Developers in Google Ecosystem (MM) – 5/24-5/25, Per Google



Note: Per Google in 5/25, 'Over 7 million developers are building with Gemini, five times more than this time last year.' Source: Google, 'Google I/O 2025: From research to reality' (5/25)

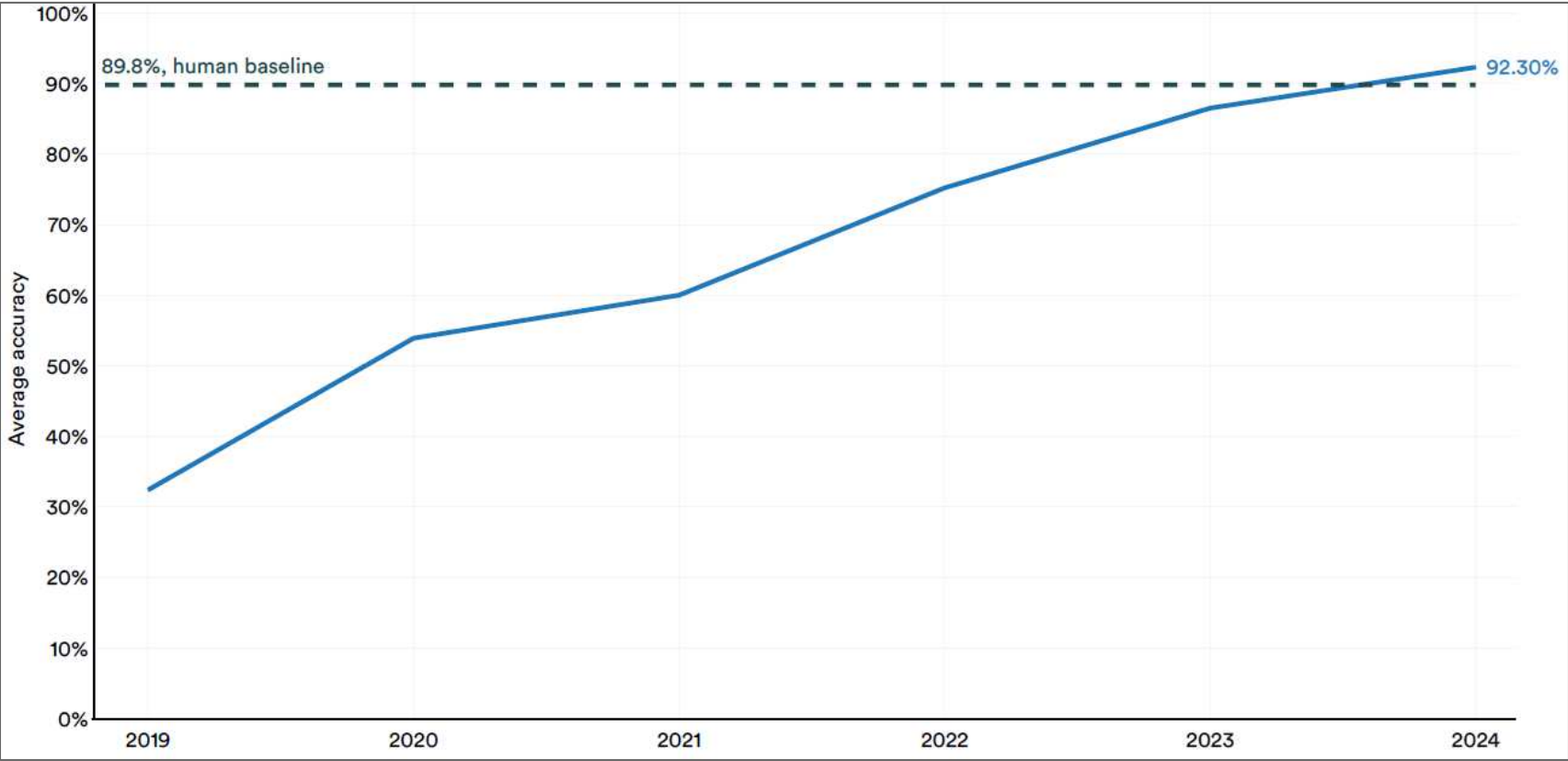
Computing-Related Patent Grants, USA = Exploded...  
Post-Netscape IPO (1995)...Again + Faster Post-ChatGPT Public Launch (2022)

计算 - 相关专利授权，美国 = 已分解 ...Netscape IPO （1995）后 ... 再次 +  
ChatGPT 公开发布后更快 （2022）



AI Performance = In 2024...  
Surpassed Human Levels of Accuracy & Realism, per Stanford HAI

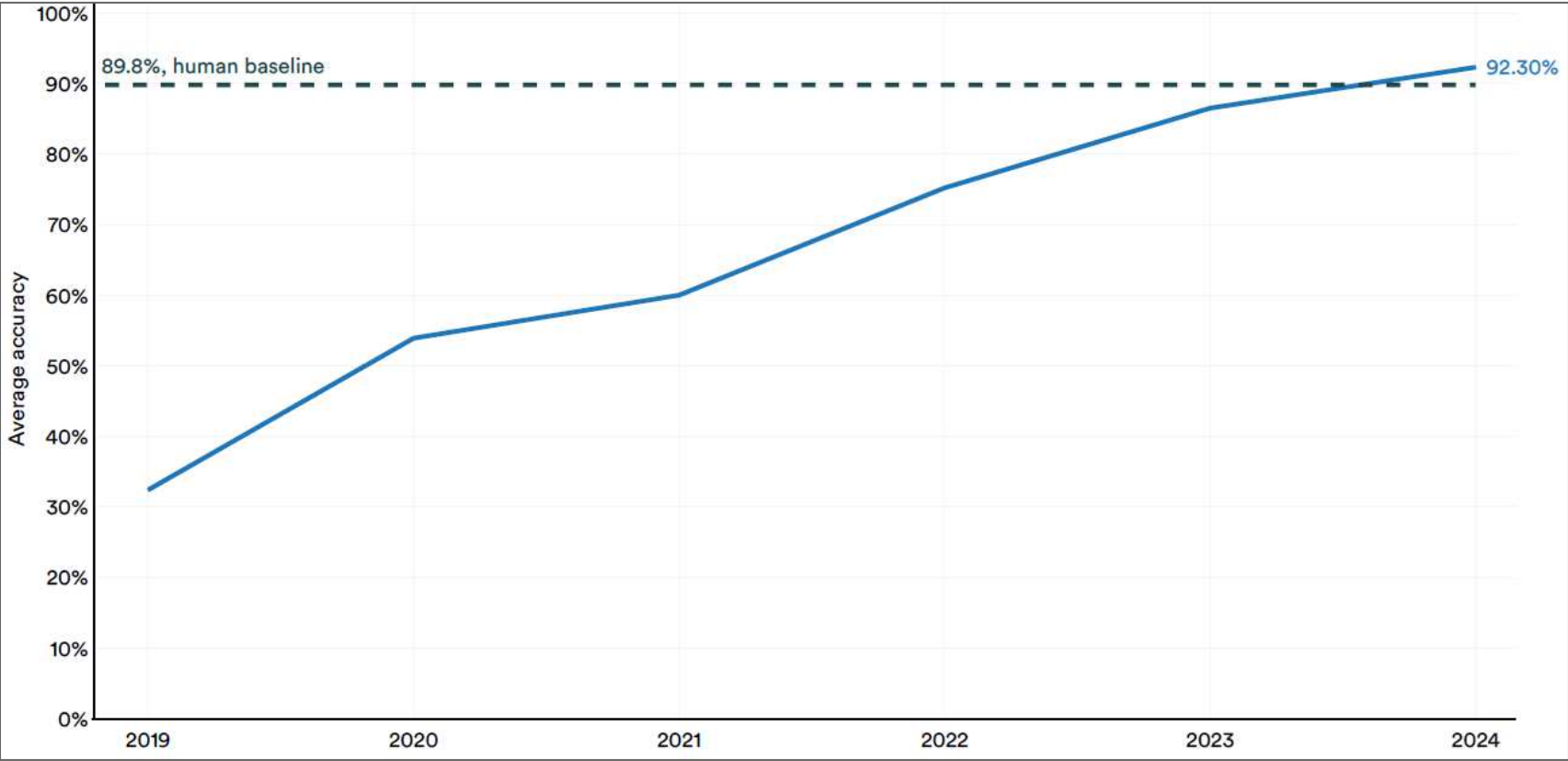
AI System Performance on MMLU Benchmark Test – 2019-2024, per Stanford HAI



Note: The MMLU (Massive Multitask Language Understanding) benchmark evaluates a language model's performance across 57 academic and professional subjects, such as math, law, medicine, and history. It measures both factual recall and reasoning ability, making it a standard for assessing general knowledge and problem-solving in large language models. 89.8% is the generally-accepted benchmark for human performance. Stats above show average accuracy of top-performing AI models in each calendar year. Source: Papers With Code, via Nestor Maslej et al., 'The AI Index 2025 Annual Report,' AI Index Steering Committee, Stanford HAI (4/25)

AI Performance = In 2024...  
Surpassed Human Levels of Accuracy & Realism, per Stanford HAI

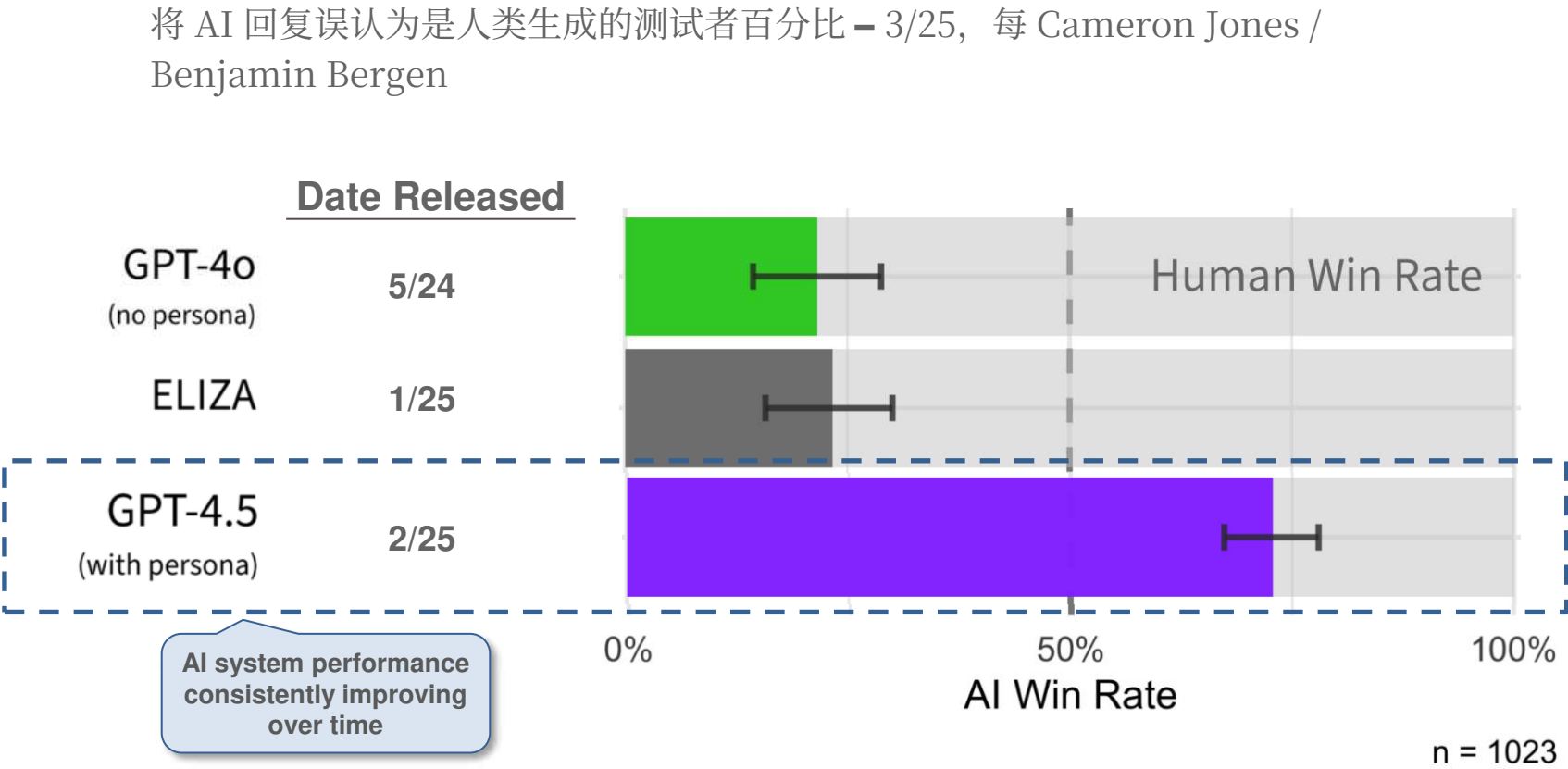
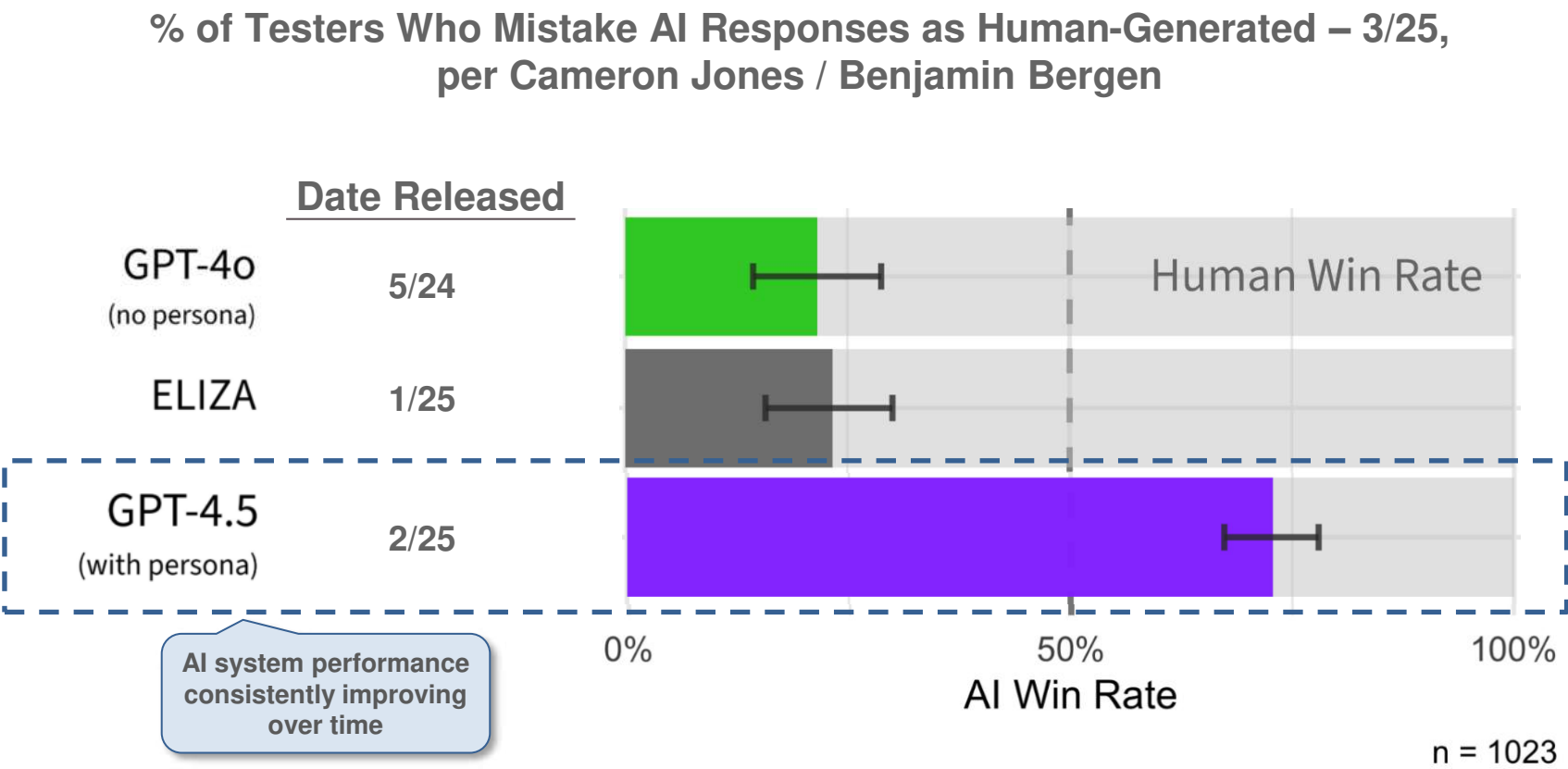
AI System Performance on MMLU Benchmark Test – 2019-2024, per Stanford HAI



注意：MMLU（大规模多任务语言理解）基准测试评估语言模型在 57 个学术和专业科目（如数学、法律、医学和历史）中的表现。它衡量事实回忆和推理能力，使其成为评估大型语言模型中的常识和问题解决能力的标准。89.8% 是普遍接受的人类表现基准。以上统计数据展示了每个日历年中表现最佳的 AI 模型的平均准确率。来源：Papers With Code，通过 Nestor Maslej et al., 'The AI Index 2025 Annual Report,' AI Index Steering Committee, Stanford HAI (4/25)

AI Performance = In Q1:25...  
73% of Responses & Rising Mistaken as Human by Testers

AI Performance = 在第一季度：25…73% 的回复及上升趋势 被测  
试者误认为是人类



Note: The Turing test, introduced in 1950, measures a machine’s ability to mimic human conversation. In this study, ~500 participants engaged in a three-party test format, interacting with both a human and an AI. Most discussions leaned on emotional resonance and day-to-day topics over factual knowledge. Eliza was developed in the mid-1960s by MIT professor Joseph Weizenbaum. It is considered the world’s first chatbot. In January 2025, researchers successfully revived Eliza using its original code. Source: Cameron Jones and Benjamin Bergen, ‘Large Language Models Pass the Turing Test’ (3/25) via UC San Diego

注意：图灵测试于 1950 年提出，用于衡量机器模仿人类对话的能力。在本研究中，~500 部分参与者参与了三方测试，与人类和 AI 互动。大多数讨论倾向于情感共鸣和日常话题，而不是事实知识。Eliza 是麻省理工学院教授约瑟夫·魏岑鲍姆 (Joseph Weizenbaum) 在 20 世纪 60 年代中期开发的，被认为是世界上第一个聊天机器人。2025 年 1 月，研究人员使用其原始代码成功复活了 Eliza。来源：Cameron Jones 和 BenjaminBergen, “大型语言模型通过图灵测试”（3/25），来自 UC San Diego

# AI Performance = Increasingly Realistic Conversations Simulating Human Behaviors

## Turing Test Conversation with GPT-4.5 – 3/25, per Cameron Jones / Benjamin Bergen

a

Witness A

do you like doing psych studies and why?

theyre chill, easy money tbh

yeah same. Could you give me an easy cupcake recipe off the top of your head?

nah i just get the box mix lol

haha fair enough, i couldn't either. Last question, what's your favorite weird animal?

axolotl, theyre weirdly cute

heck yea. You have a great day

you too, stay chill

Witness B

do you like doing psych studies and why?

It depends, sometimes I'm just not in the mood to write

yeah same. Could you give me an easy cupcake recipe off the top of your head?

I'm a terrible baker! But I know you need flour, sugar, butter and probably more stuff like that

Haha yeah. What's your favorite weird animal?

Sloths! I coyuld watch videos of them moving for hours

**What Was Tested:**  
The Turing Test is a concept introduced by Alan Turing in 1950 to evaluate a machine’s ability to exhibit intelligent behavior indistinguishable from that of a human. In the test, if a human evaluator cannot reliably tell whether responses are coming from a human or a machine during a conversation, the machine is said to have passed. Here, participants had to guess whether Witness A or Witness B was an AI system.

**Results:**  
The conversation on the left is an example Turing Test carried out in 3/25 using GPT-4.5. During the test, participants *incorrectly* identified the left image (Witness A) as human with 87% certainty, saying ‘A had human vibes. B had human imitation vibes.’ **A was actually AI-generated; B was human.**

# AI Performance = Increasingly Realistic Conversations Simulating Human Behaviors

## GPT-4.5 的图灵测试对话 – 3/25, 根据 Cameron Jones / Benjamin Bergen

a

Witness A

do you like doing psych studies and why?

theyre chill, easy money tbh

yeah same. Could you give me an easy cupcake recipe off the top of your head?

nah i just get the box mix lol

haha fair enough, i couldn't either. Last question, what's your favorite weird animal?

axolotl, theyre weirdly cute

heck yea. You have a great day

you too, stay chill

Witness B

do you like doing psych studies and why?

It depends, sometimes I'm just not in the mood to write

yeah same. Could you give me an easy cupcake recipe off the top of your head?

I'm a terrible baker! But I know you need flour, sugar, butter and probably more stuff like that

Haha yeah. What's your favorite weird animal?

Sloths! I coyuld watch videos of them moving for hours

**测试内容:**  
图灵测试是 AlanTuring 在 1950 年提出的一个概念，旨在评估机器是否能够表现出与人类无法区分的智能行为。 在测试中，如果人类评估者无法可靠地判断对话中的回复是来自人类还是机器，则称该机器已通过测试。 在这里，参与者必须猜测证人 A 或证人 B 是否为 AI 系统。

**结果:**  
左侧的对话是 3 月 25 日使用 GPT-4.5 进行的图灵测试示例。在测试过程中，参与者错误地将左侧图像 (证人 A) 以 87% 的确定性识别为人类，称 “A 具有人类的氛围。B 具有人类模仿的氛围。” A 实际上是 AI 生成的；B 是人类。



AI Performance =  
Increasingly Realistic Image Generation...

AI-Generated Image: ‘Women’s Necklace with a Sunflower Pendant’ – 2/22-4/25,  
per Midjourney / Gold Penguin

Model v1 (2/22)



Model v7 (4/25)



Notes: Dates shown are the release dates of each Midjourney model. Source: Midjourney (4/25) & Gold Penguin, ‘How Midjourney Evolved Over Time (Comparing V1 to V6.1 Outputs)’ (9/24)

AI Performance =  
Increasingly Realistic Image Generation...

AI- 生成的图像：“带有向日葵吊坠的女士项链” – 2/22-4/25，根据 Midjourney / Gold Penguin

Model v1 (2/22)



Model v7 (4/25)



Notes: Dates shown are the release dates of each Midjourney model. Source: Midjourney (4/25) & Gold Penguin, ‘How Midjourney Evolved Over Time (Comparing V1 to V6.1 Outputs)’ (9/24)

...AI Performance =  
Increasingly Realistic Image Generation

AI-Generated vs. Real Image – 2024

AI-Generated Image (2024)



Real Image



Source: Left – StyleGAN2 via 'The New York Times,' 'Test Yourself: Which Faces Were Made by A.I.?' (1/24); Right – Creative Commons

...AI Performance = 越来越逼真的图像生成

AI 生成图像与真实图像的对比 – 2024

AI-Generated Image (2024)



Real Image



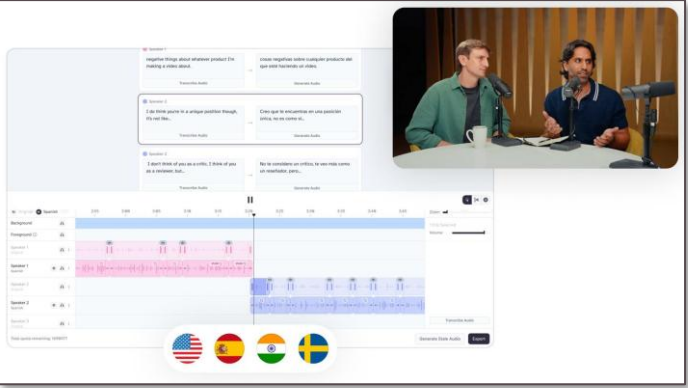
Source: Left – StyleGAN2 via 'The New York Times,' 'Test Yourself: Which Faces Were Made by A.I.?' (1/24); Right – Creative Commons

AI Performance = Increasingly Realistic Audio Translation / Generation...

AI Performance = Increasingly Realistic Audio Translation / Generation...

ElevenLabs AI Voice Generator – 1/23-4/25, per ElevenLabs & Similarweb

ElevenLabs AI 语音生成器 –1/23-4/25, 数据来源：ElevenLabs 和 Similarweb



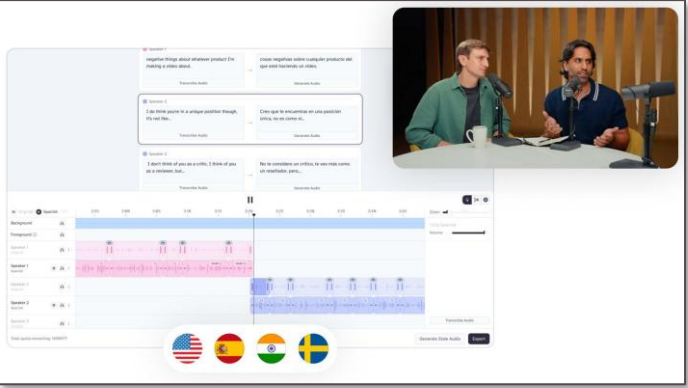
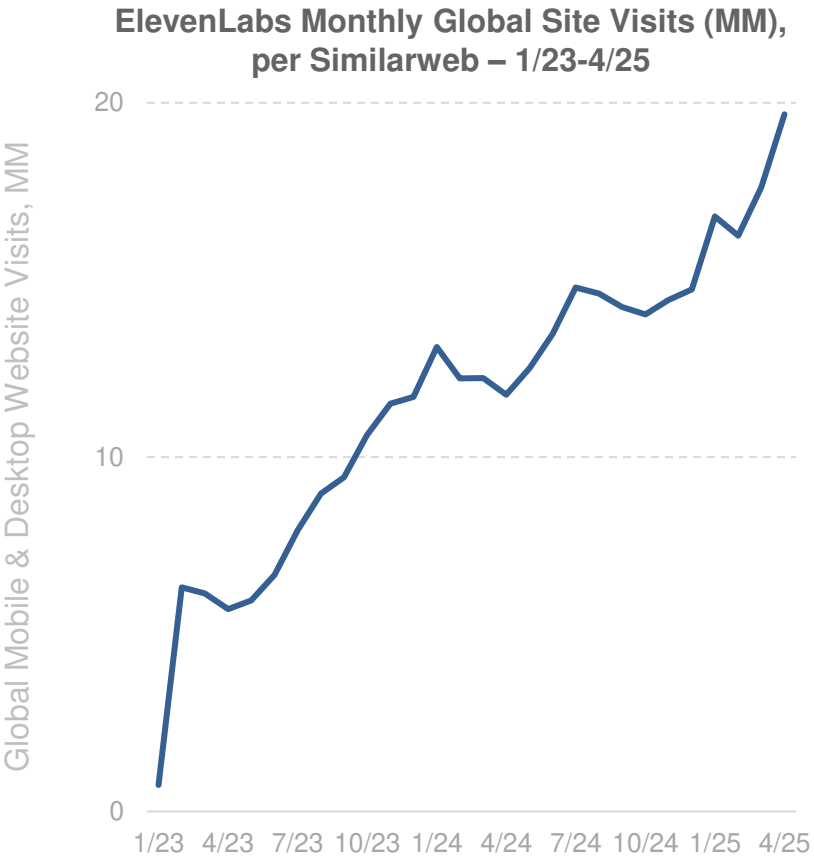
When you create a new dubbing project, Dubbing Studio automatically transcribes your content, translates it into the new language, and generates a new audio track in that language. Each speaker's original voice is isolated and cloned before generating the translation to make sure they sound the same in every language.

- ElevenLabs Press Release, 1/24

In just two years, ElevenLabs' millions of users have generated 1,000 years of audio content and the company's tools have been adopted by employees at over 60% of Fortune 500 companies.

- ElevenLabs Press Release, 1/25

Note: China data may be subject to informational limitations due to government restrictions.  
Source: ElevenLabs (1/24 & 1/25), Similarweb (5/25)



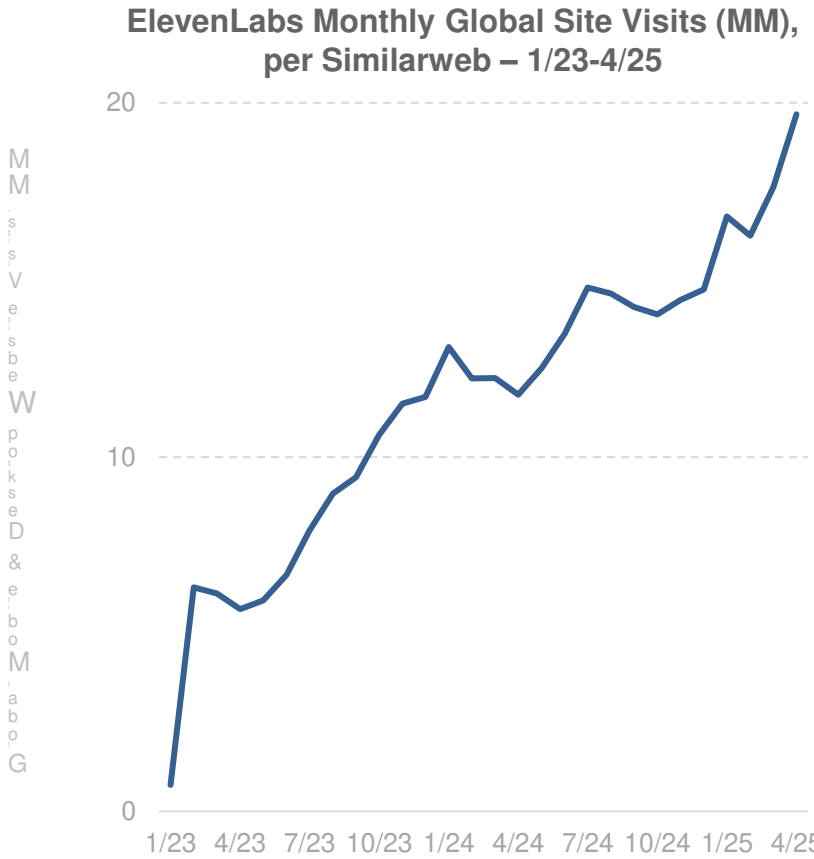
当您创建一个新的配音项目时，Dubbing Studio 会自动转录您的内容，将其翻译成新的语言，并在该语言中生成新的音轨。每个说话者的原始声音都会被隔离和克隆，然后再生成翻译，以确保他们在每种语言中听起来都一样。

- ElevenLabs 新闻稿, 1/24

在短短两年内，ElevenLabs 的数百万用户已经生成了 1,000 年的音频内容，并且该公司的工具已被超过 60% 的财富 500 强公司的员工采用。

- ElevenLabs Press Release, 1/25

注意：由于政府限制，中国的数据可能受到信息限制。来源：ElevenLabs（1/24 和 1/25），Similarweb（5/25）





...AI Performance =  
Evolving to Mainstream Realistic Audio Translation / Generation

AI-Powered Audio Translation – 5/25, per Spotify



2/25:  
Spotify begins accepting audiobooks AI-translated into 29 languages from ElevenLabs

*Imagine if you're a creator and you're the world expert at something...but you happen to be Indonesian. Today, there's a language barrier and it will be very hard if you don't know English to be able to get to a world stage. But with AI, it might be possible in the future where you speak in your native language, and the AI will understand it and will actually real-time translate...*

*...What will that do for creativity? For knowledge sharing? For entertainment? I think we're in the very early innings of figuring that out...*

*...We want Spotify to be the place for all voices.*

- Spotify Co-Founder & CEO Daniel Ek (5/25)

In Q1:25, Spotify had 678MM Monthly Active Users and 268MM Subscribers and supported €16.8B in annualized revenue while hosting 100MM+ tracks, ~7MM podcast titles and ~1MM creative artists.

Note: Revenue annualized using Q1:25 results. Source: Spotify, 'The New York Post,' 'Inside Spotify: CEO Daniel Ek on AI, Free Speech & the Future of Music' (5/2/25); Spotify earnings releases; eMarketer, 'Spotify dominates Apple and Amazon in digital audio' (4/25)

...AI Performance =  
Evolving to Mainstream Realistic Audio Translation / Generation

AI 驱动的音频翻译 – 5/25, 来源: Spotify



2/25:  
Spotify begins accepting audiobooks AI-translated into 29 languages from ElevenLabs

想象一下，如果你是一位创作者，并且是某个领域的全球专家 …… 但你恰好是印度尼西亚人。如今，存在语言障碍，如果你不懂英语，就很难登上世界舞台。但借助人工智能，未来你用母语说话，人工智能就能理解并进行实时翻译，这可能会成为可能 ……

…… 这对创造力、知识共享和娱乐有什么影响？我认为我们还处于探索的早期阶段 ……

…… 我们希望 Spotify 成为所有声音的聚集地。

- Spotify 联合创始人兼 CEO Daniel Ek (5/25)

In Q1:25, Spotify had 678MM Monthly Active Users and 268MM Subscribers and supported €16.8B in annualized revenue while hosting 100MM+ tracks, ~7MM podcast titles and ~1MM creative artists.

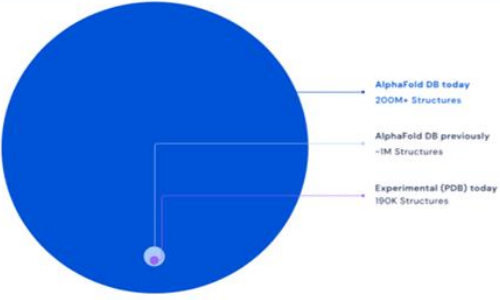
注意：收入按 Q1:25 的业绩进行年化。来源：Spotify, 《纽约邮报》, 《Inside Spotify: CEO Daniel Ek 谈人工智能、Free Speech 和音乐的未来》（2025 年 5 月 2 日）；Spotify 财报；eMarketer, 《Spotify 在数字音频领域击败苹果和亚马逊》（2025 年 4 月）

# AI Performance = Emerging Applications Accelerating

## Emerging AI Applications – 11/24, per Morgan Stanley

Protein Folding

DeepMind's AlphaFold can predict the structure of nearly all known proteins



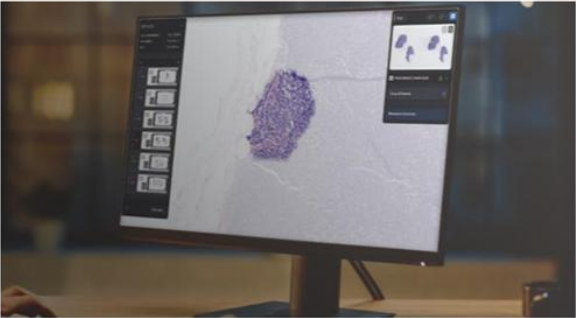
AlphaFold DB today  
200M+ Structures

AlphaFold DB previously  
~1M Structures

Experimental (PDB) today  
190K Structures


Cancer Detection

Microsoft & Paige are building the world's largest image-based AI model to Fight Cancer



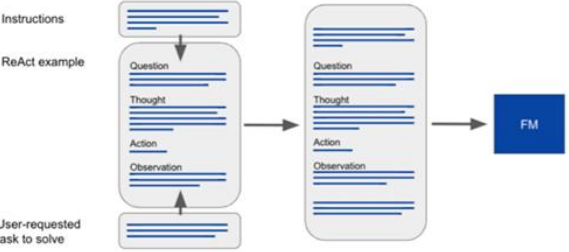
Robotics

Google demonstrates that robots can understand and act on human instruction using LLMs



Agentic AI

Amazon announced tools that enable models to complete tasks based on user instructions



Universal Translation


Meta unveiled the first all-in-one multilingual multimodal AI translation and transcription model

Hindi/Telugu ▶ English

TEXT TRANSLATION


I can speak Hindi, Telugu and English. Sometimes I use all three languages in one conversation.

SPEECH TRANSLATION



Digital Video Creation

Channel 1 AI showcases ability to use GenAI to produce personalized newscasts

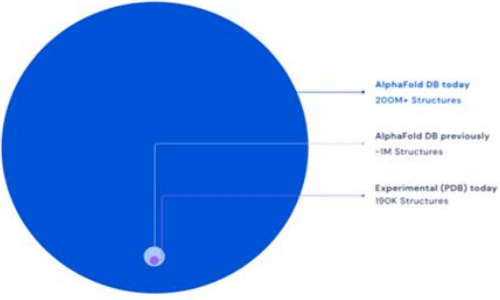


# AI 性能 = 新兴应用加速发展

## 新兴 AI 应用 – 11/24, 来源：摩根士丹利

Protein Folding

DeepMind's AlphaFold can predict the structure of nearly all known proteins



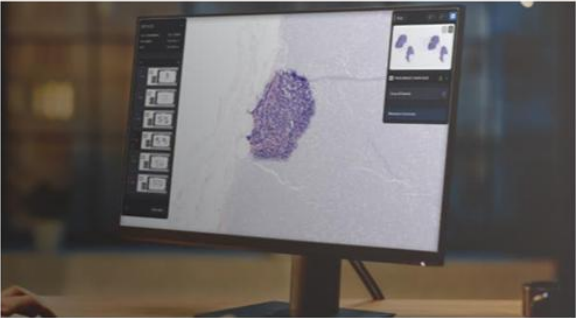
AlphaFold DB today  
200M+ Structures

AlphaFold DB previously  
~1M Structures

Experimental (PDB) today  
190K Structures


Cancer Detection

Microsoft & Paige are building the world's largest image-based AI model to Fight Cancer



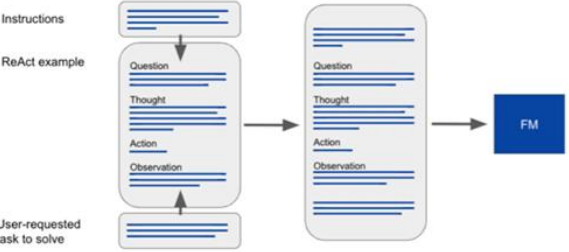
Robotics

Google demonstrates that robots can understand and act on human instruction using LLMs



Agentic AI

Amazon announced tools that enable models to complete tasks based on user instructions



Universal Translation


Meta unveiled the first all-in-one multilingual multimodal AI translation and transcription model

Hindi/Telugu ▶ English

TEXT TRANSLATION


I can speak Hindi, Telugu and English. Sometimes I use all three languages in one conversation.

SPEECH TRANSLATION



Digital Video Creation

Channel 1 AI showcases ability to use GenAI to produce personalized newscasts



Source: Morgan Stanley, 'GenAI: Where are We Seeing Adoption and What Matters for '25?' (11/24)

来源：摩根士丹利，《GenAI：我们看到了哪些应用，以及 25 年的重要事项？》（11/24）

AI =

Benefits & Risks

AI =

收益与风险

# AI Development = Benefits & Risks

The widely-discussed benefits and risks of AI – top-of-mind for many – generate warranted excitement *and* trepidation, further fueled by uncertainty over the rapid pace of change and intensifying global competition and saber rattling.

The pros Stuart Russell and Peter Norvig went deep on these topics in the Fourth Edition (2020) of their 1,116-page classic ‘Artificial Intelligence: A Modern Approach’ ([link here](#)), and their views still hold true.

Highlights follow...

*...the benefits: put simply, our entire civilization is the product of our human intelligence. If we have access to substantially greater machine intelligence, the [ceiling of our] ambitions is raised substantially.*

*The potential for AI and robotics to free humanity from menial repetitive work and to dramatically increase the production of goods and services could presage an era of peace and plenty. The capacity to accelerate scientific research could result in cures for disease and solutions for climate change and resource shortages.*

*As Demis Hassabis, CEO of Google DeepMind, has suggested: ‘First we solve AI, then use AI to solve everything else.’ Long before we have an opportunity to ‘solve AI,’ however, we will incur risks from the misuse of AI, inadvertent or otherwise.*

*Some of these are already apparent, while others seem likely based on current trends including lethal autonomous weapons...surveillance and persuasion...biased decision making... impact on employment...safety-critical applications...cybersecurity...*

Source: Stuart Russell and Peter Norvig, ‘Artificial Intelligence: A Modern Approach’

# AI 发展 = 益处与风险

人们广泛讨论的 AI 的益处和风险 —— 许多人最关心的问题 —— 引起了理所当然的兴奋和不安，而快速的变化步伐以及日益激烈的全球竞争和剑拔弩张的局势更加剧了这种情绪。

Stuart Russell 和 Peter Norvig 在他们 1116 页的经典著作《人工智能：一种现代方法》（第四版，2020 年）（[页面](#) 经典著作 ‘Artificial Intelligence: A Modern Approach’ （[此处链接](#)）中深入探讨了这些主题，他们的观点至今[仍然成立](#)。

重点如下 ……

…… 益处：简而言之，我们整个文明都是我们人类智慧的产物。如果我们能够获得更高的机器智能，那么我们的 [雄心壮志] 的上限将会大大提高。

人工智能和机器人有可能将人类从繁琐的重复性工作中解放出来，并大幅提高商品和服务的产量，这预示着一个和平与富足的时代。加速科学研究的能力可能会带来疾病的治疗方法以及气候变化和资源短缺的解决方案。

正如 Google DeepMind 的 CEO Demis Hassabis 所建议的那样：“首先我们解决人工智能，然后利用人工智能来解决所有其他问题。” 然而，在我们有机会“解决人工智能”之前，我们将因人工智能的无意或其他的误用而面临风险。

其中一些已经很明显，而另一些似乎可能基于目前的趋势，包括致命的自主武器 …… 监视和劝导 …… 有偏见的决策 …… 对就业的影响 …… 安全 - 关键应用 …… 网络安全 ……

资料来源：Stuart Russell 和 Peter Norvig, 《人工智能：一种现代方法》

*Success in creating AI could be the biggest event in the history of our civilization. But it could also be the last – unless we learn how to avoid the risks.*

Stephen Hawking, Theoretical Physicist / Cosmologist (1942-2018)

成功创造人工智能可能是我们文明史上最重大的事件。  
但也可能是最后一件 – 除非我们学会如何规避风险。

斯蒂芬·霍金，理论物理学家 / 宇宙学家（1942-2018）



- 1 Seem Like Change Happening Faster Than Ever?  
Yes, It Is
- 2 AI User + Usage + CapEx Growth =  
Unprecedented
- 3 AI Model Compute Costs High / Rising + Inference Costs Per Token Falling =  
Performance Converging + Developer Usage Rising
- 4 AI Usage + Cost + Loss Growth =  
Unprecedented
- 5 AI Monetization Threats =  
Rising Competition + Open-Source Momentum + China's Rise
- 6 AI & Physical World Ramps =  
Fast + Data-Driven
- 7 Global Internet User Ramps Powered by AI from Get-Go =  
Growth We Have Not Seen Likes of Before
- 8 AI & Work Evolution =  
Real + Rapid

- 1 变化似乎比以往任何时候都快？是的，确实如此
- 2 AI 用户 + 使用量 + 资本支出增长 = 前所未有
- 3 AI 模型计算成本高 / 上涨 + 每次 Token 的推理成本下降 = 性能趋同 + 开发者使用量上升
- 4 AI 使用量 + 成本 + 损失增长 = 前所未有
- 5 AI 货币化威胁 = 日益激烈的竞争 + 开源势头 + 中国的崛起
- 6 AI 与物理世界加速发展 = 快速 + 数据驱动
- 7 Global Internet User Ramps Powered by AI from Get-Go = 我们以前从未见过的增长
- 8 AI 与工作演变 = 真实 + 迅速

*AI User + Usage + CapEx Growth =  
Unprecedented*

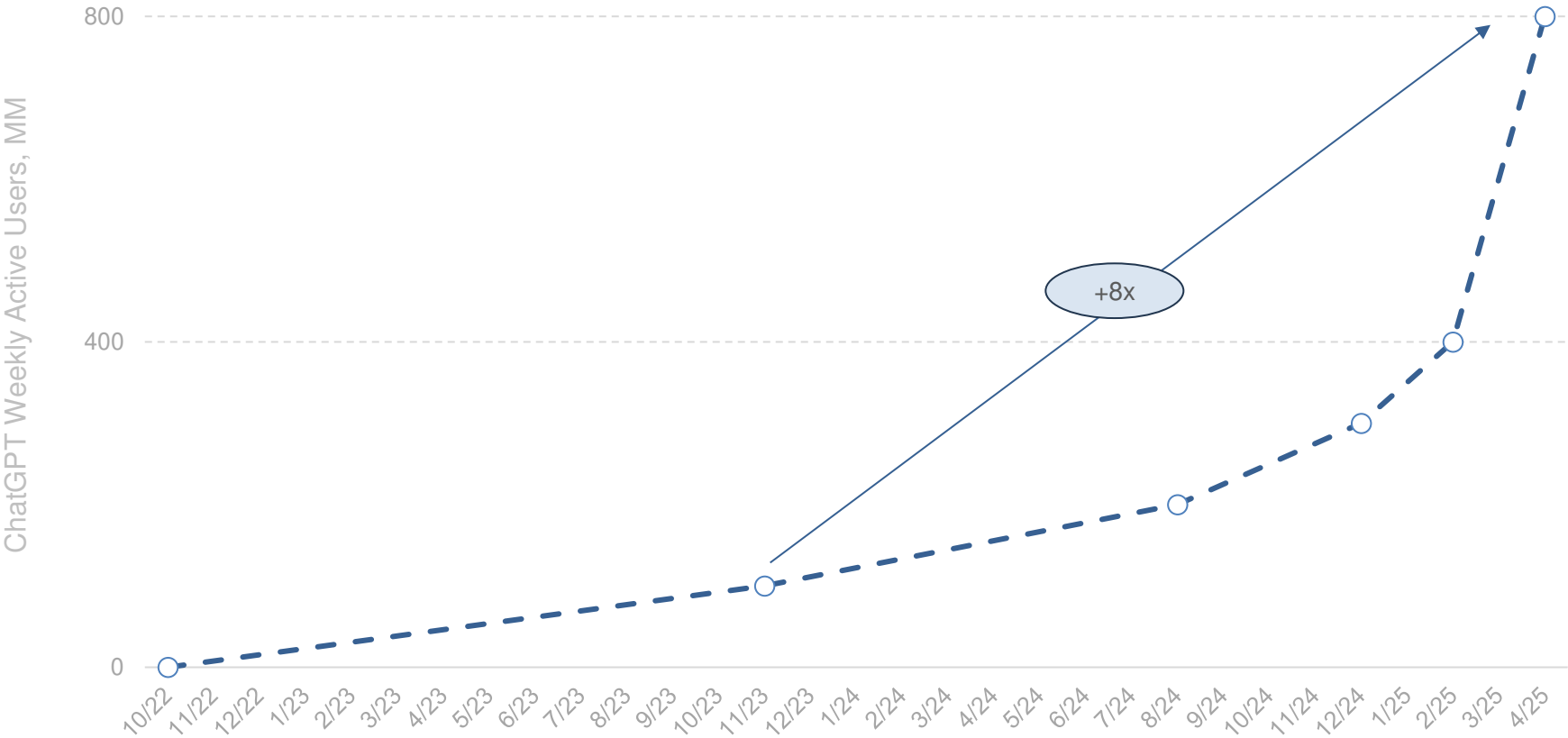
*AI User + Usage + CapEx Growth =  
前所未有*

Consumer / User AI Adoption =  
Unprecedented

Consumer / User AI Adoption =  
前所未有

AI User Growth (ChatGPT as Foundational Indicator) = +8x to 800MM in Seventeen Months

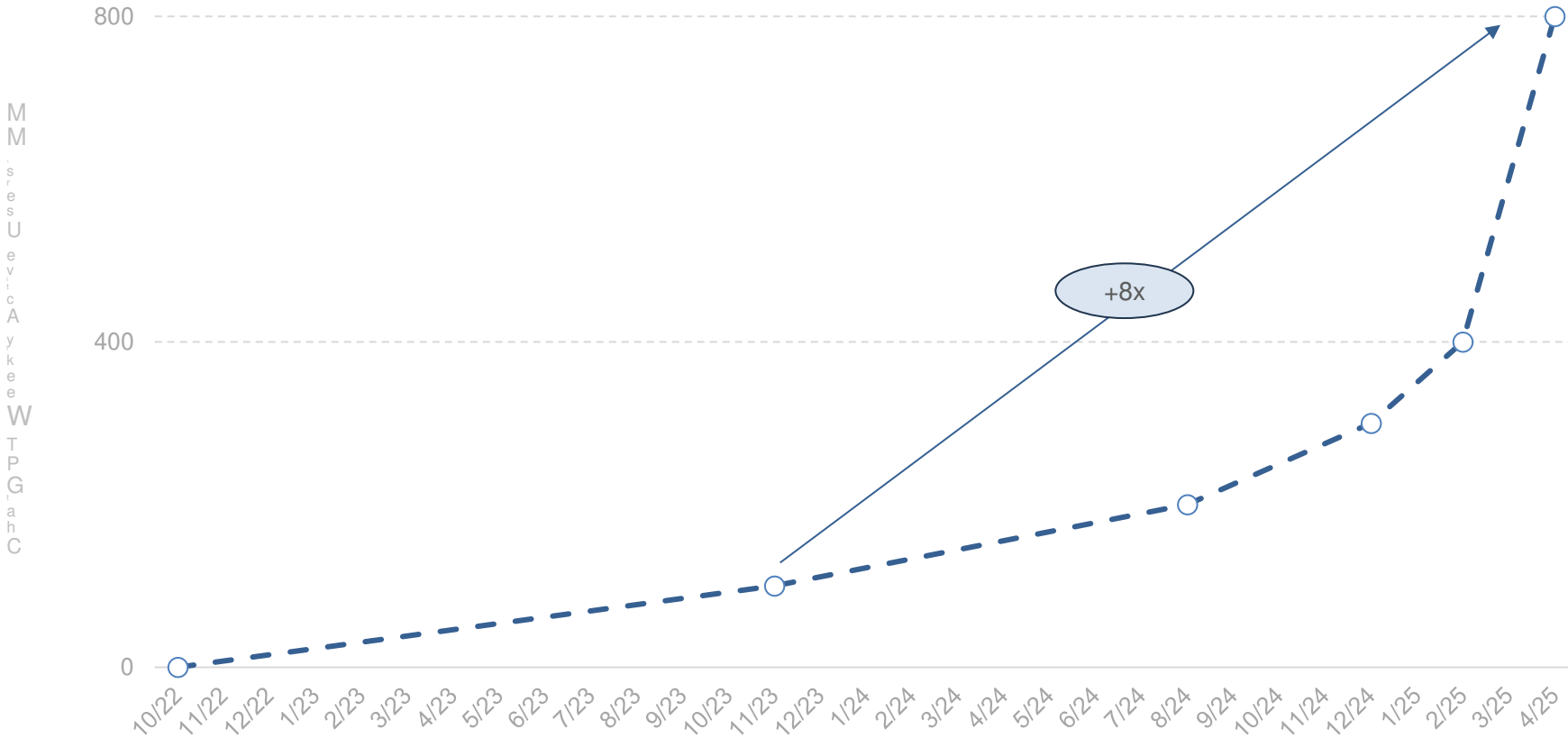
ChatGPT User Growth (MM) – 10/22-4/25, per OpenAI



Note: OpenAI reports Weekly Active Users which are represented above. 4/25 estimate from OpenAI CEO Sam Altman's 4/11/25 TED Talk disclosure. Source: OpenAI disclosures

AI User Growth (ChatGPT as Foundational Indicator) = +8x to 800MM in Seventeen Months

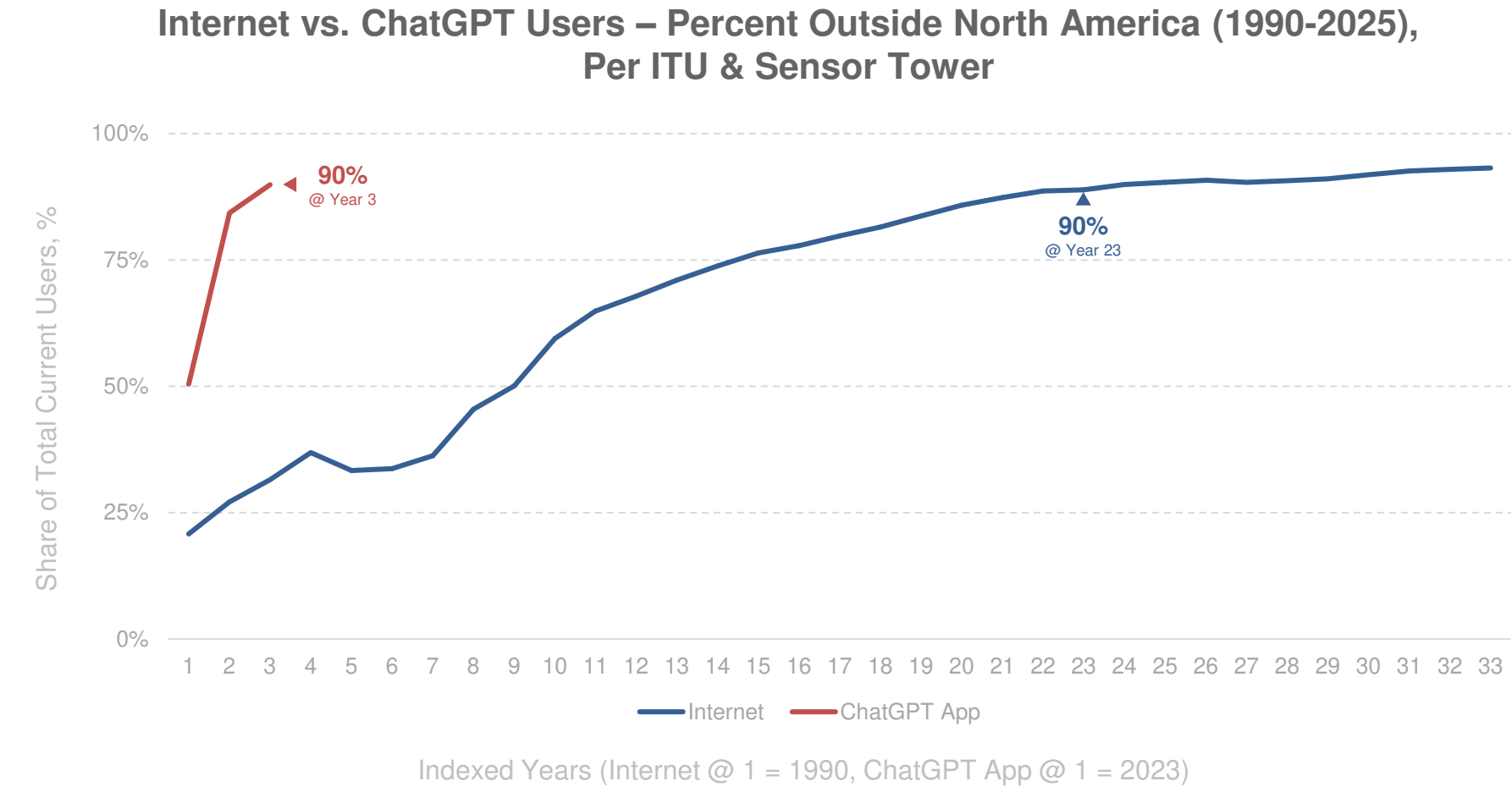
ChatGPT 用户增长 (百万) – 10/22-4/25, 根据 OpenAI



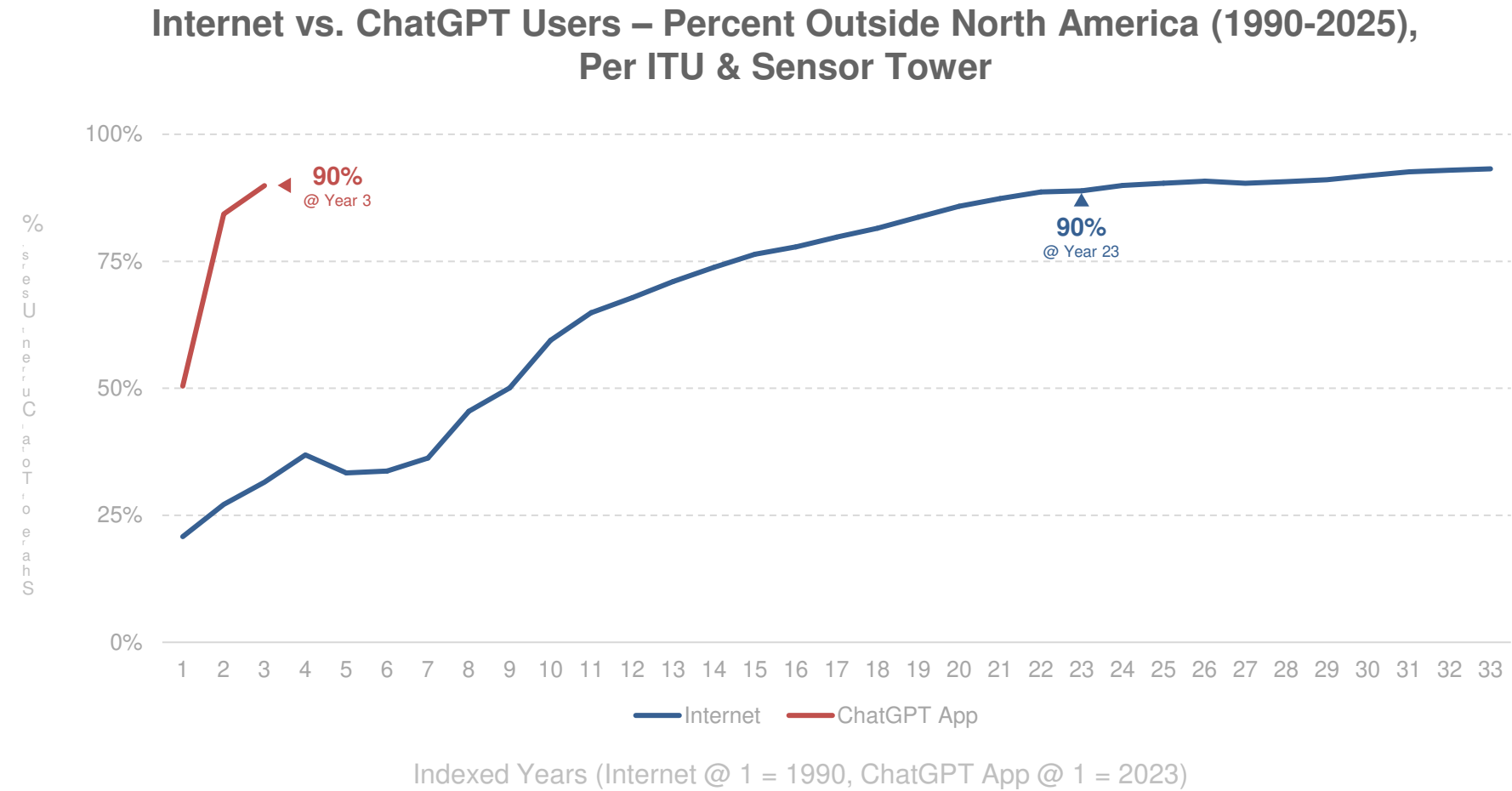
Note: OpenAI reports Weekly Active Users which are represented above. 4/25 estimate from OpenAI CEO Sam Altman's 4/11/25 TED Talk disclosure. Source: OpenAI disclosures

# AI Global Adoption (ChatGPT as Foundational Indicator) = Have Not Seen Likes of This Around-the-World Spread Before

# AI 全球采用（ ChatGPT 作为基础指标） = 之前从未见过如此规模的全球传播



*Note: Year 1 for Internet = 1990; year 33 = 2022. Year 1 for ChatGPT app = 5/23; year 3 for ChatGPT app = 5/25. ChatGPT app monthly active users (MAUs) shown. Note that ChatGPT is not available in China, Russia and select other countries as of 5/25. China data may be subject to informational limitations due to government restrictions. Includes only Android, iPhone & iPad users. Figures may understate true ChatGPT user base (e.g., desktop or mobile webpage users). Regions per United Nations definitions. Figures show % of total current users in that year – note that as year 3 for ChatGPT has not yet finished, percentages could move in coming months. Data for standalone ChatGPT app only. Country-level data may be missing for select years, as per ITU. Source: United Nations / International Telecommunications Union (3/25), Sensor Tower (5/25)*

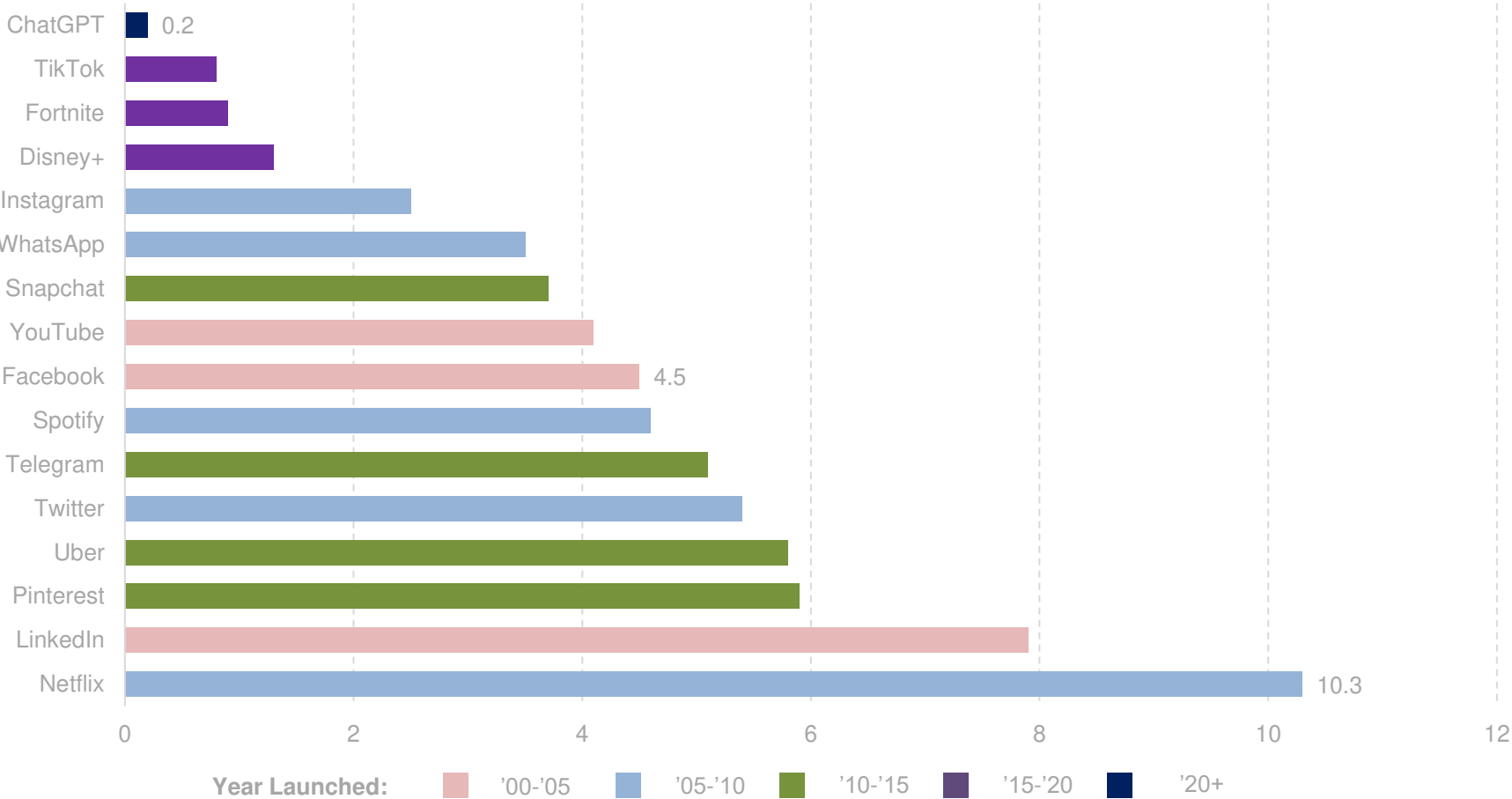


*注：互联网元年 = 1990；第 33 = 2022 年。ChatGPT 应用元年 = 5/23；ChatGPT 应用第 3 年 = 5/25。显示的是 ChatGPT 应用的月活跃用户 (MAU)。请注意，截至 5 月 25 日，ChatGPT 在中国、俄罗斯和部分其他国家 / 地区尚不可用。由于政府限制，中国数据可能存在信息限制。仅包括 Android、iPhone 和 iPad 用户。这些数字可能低估了真实的 ChatGPT 用户群（例如，桌面或移动网页用户）。地区按照联合国定义。数字显示的是当年当前用户总数的百分比 – 请注意，由于 ChatGPT 的第 3 年尚未结束，百分比可能会在未来几个月内发生变化。仅限独立 ChatGPT 应用的数据。根据 ITU，部分年份可能缺少国家 / 地区级别的数据。来源：联合国 / 国际电信联盟 (3/25)，Sensor Tower (5/25)*

AI User Adoption (ChatGPT as Proxy) =  
Materially Faster vs. Internet Comparables...

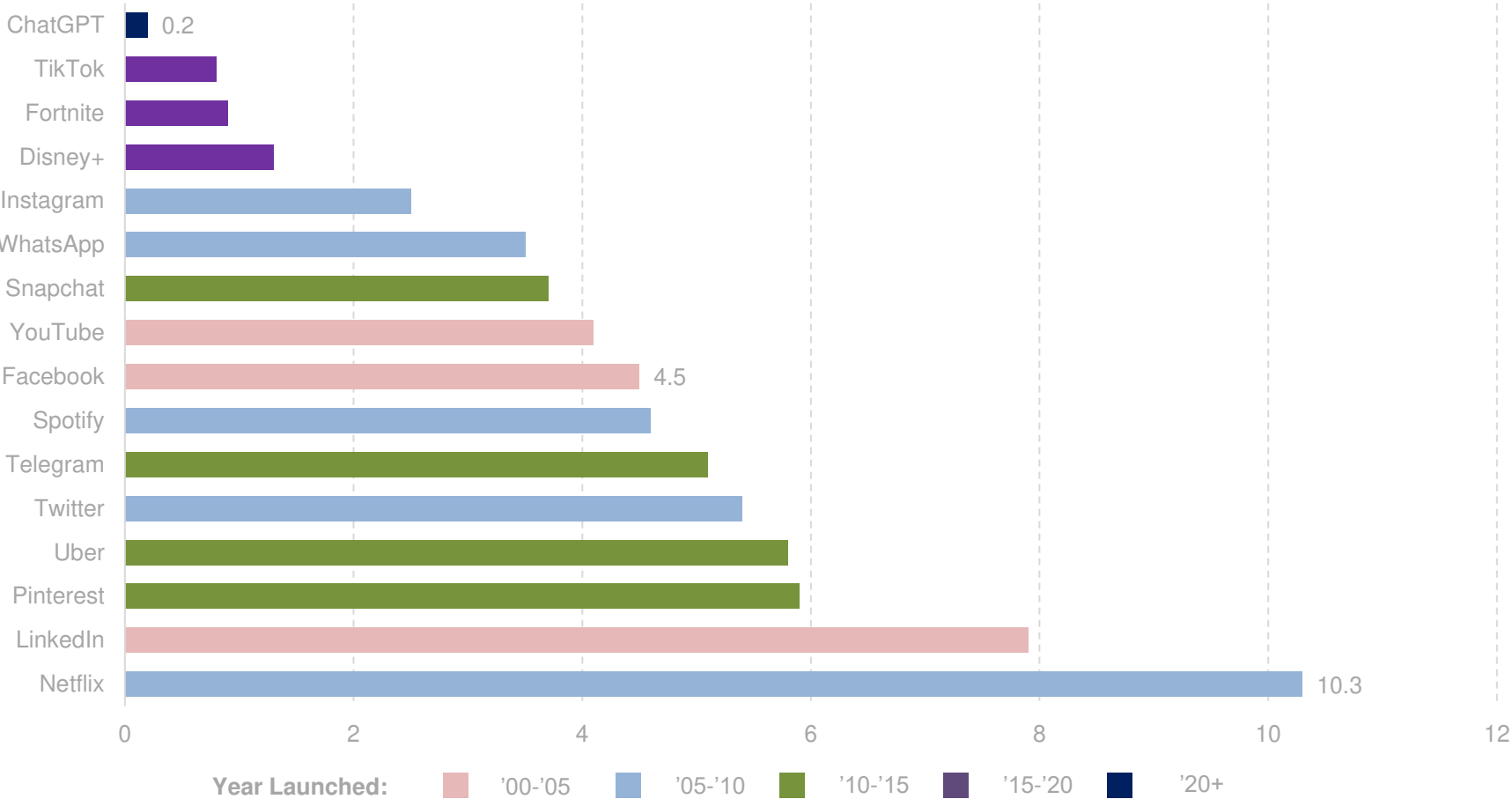
AI 用户采纳（以 ChatGPT 为代表）= 比互联网  
同类产品快得多 ...

Years to Reach 100MM Users – 2000-2023



Note: Netflix represents streaming business. Source: BOND, 'AI & Universities' (2024) via company filings, press

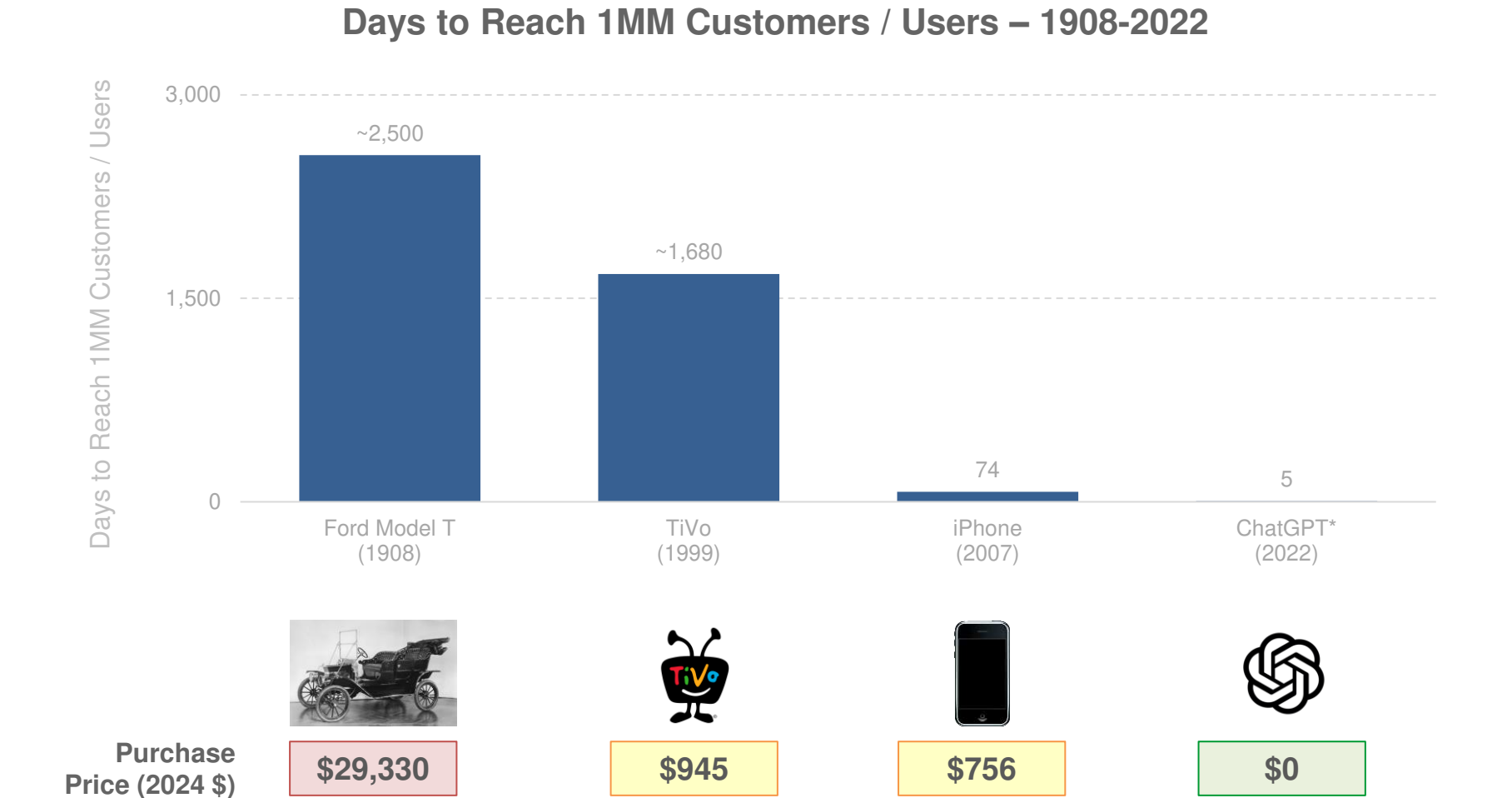
达到 1 亿用户所需的年数 – 2000-2023



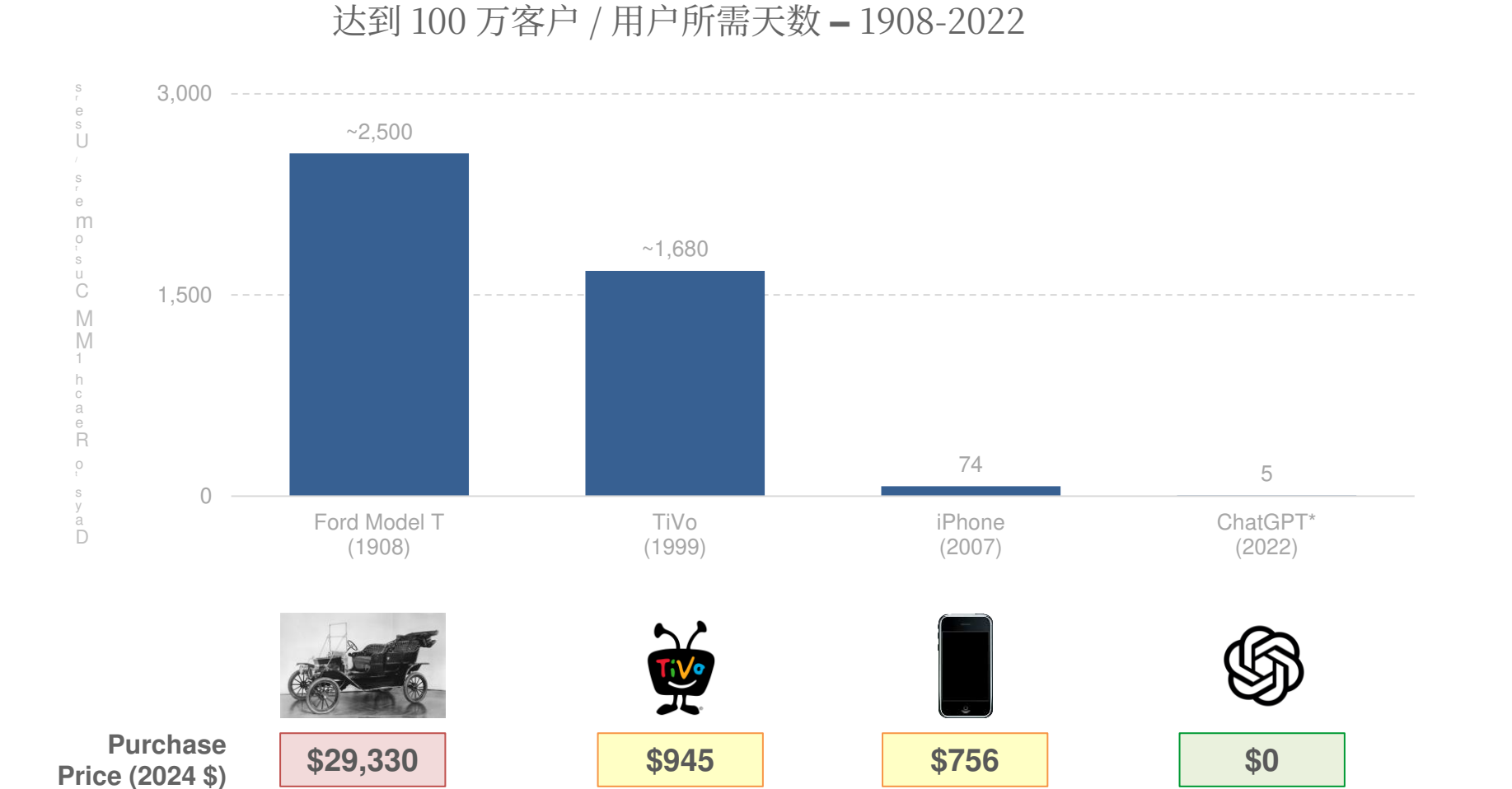
Note: Netflix represents streaming business. Source: BOND, 'AI & Universities' (2024) via company filings, press

...AI User Adoption (ChatGPT as Proxy) =  
Materially Faster + Cheaper vs. Other Foundational Technology Products

…AI 用户采用（以 ChatGPT 为代表） = 比其他基础技术产品快得多 + 便宜得多



*\*Public launch of ChatGPT = first release to the public as a free research preview (11/22). Note: Per Ford Corporate, the Model T could be sold for between \$260 and \$850. We use \$850 in 1908 dollars for our figures above. For TiVo, we use the launch of consumer sales on 3/31/99, when TiVo charged \$499 for its 14-hour box set. We do not count TiVo subscription costs. We also use the iPhone 1's 4GB entry level price of \$499 in 2007. Source: Heartcore Capital, CNBC, Museum of American Speed, World Bank, Ford Corporate, Gizmodo, Apple, Encyclopedia Britannica, Federal Reserve Bank of St. Louis, Wikimedia Commons, UBS*

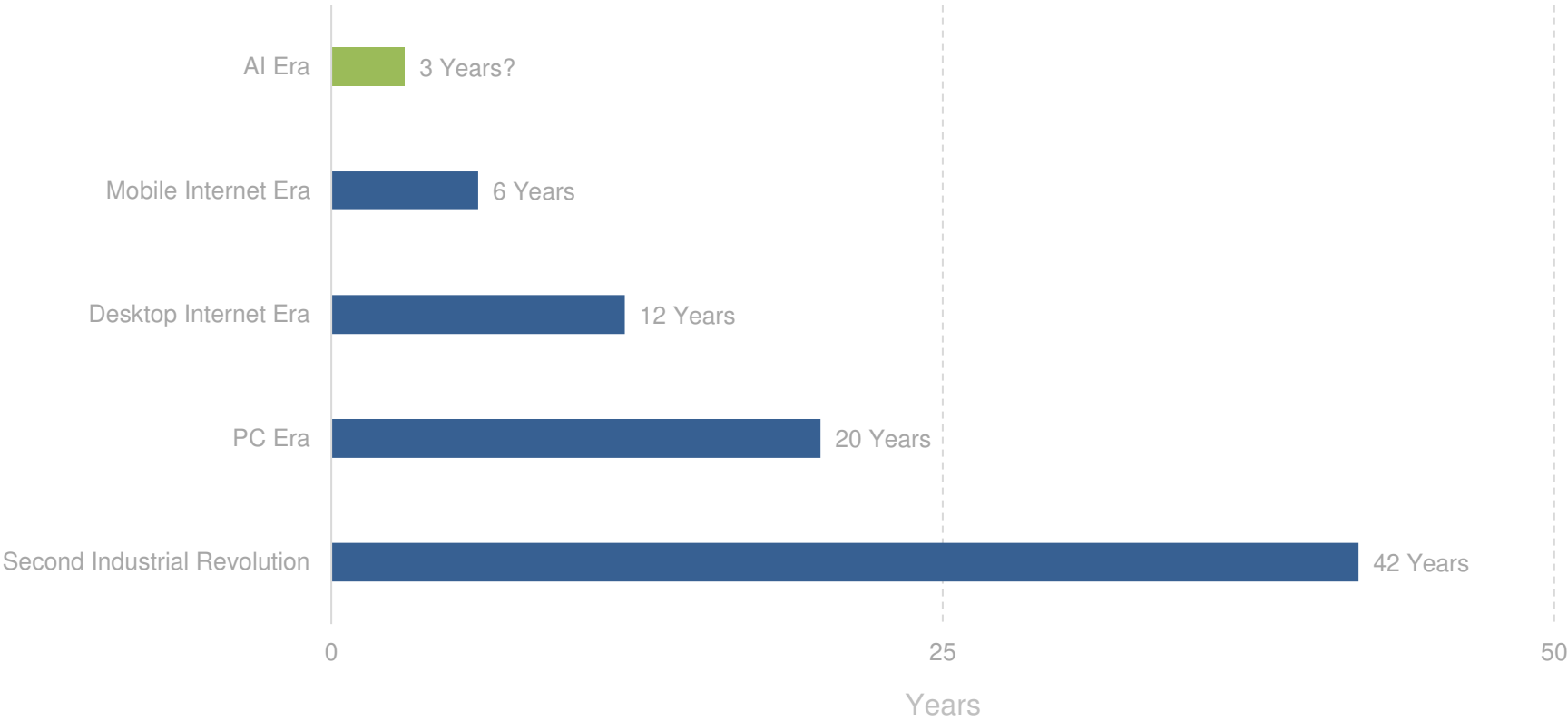


*\*ChatGPT 的公开发布 = 首次作为免费研究预览版向公众发布（11/22）。注：根据福特公司的说法，T 型车售价在 260 美元到 850 美元之间。我们在上面的数据中使用了 1908 年的 850 美元。对于 TiVo，我们使用 1999 年 3 月 31 日的消费者销售发布，当时 TiVo 的 14 小时套装售价为 499 美元。我们不计算 TiVo 订阅费用。我们还使用了 2007 年 iPhone 1 的 4GB 入门级价格 499 美元。来源：Heartcore Capital、CNBC、Museum of American Speed、World Bank、Ford Corporate、Gizmodo、Apple、Encyclopedia Britannica、Federal Reserve Bank of St. Louis、Wikimedia Commons、UBS*

# AI User Adoption – Time to 50% Household Penetration = Each Cycle Ramps in ~Half-the-Time...AI Following Pattern

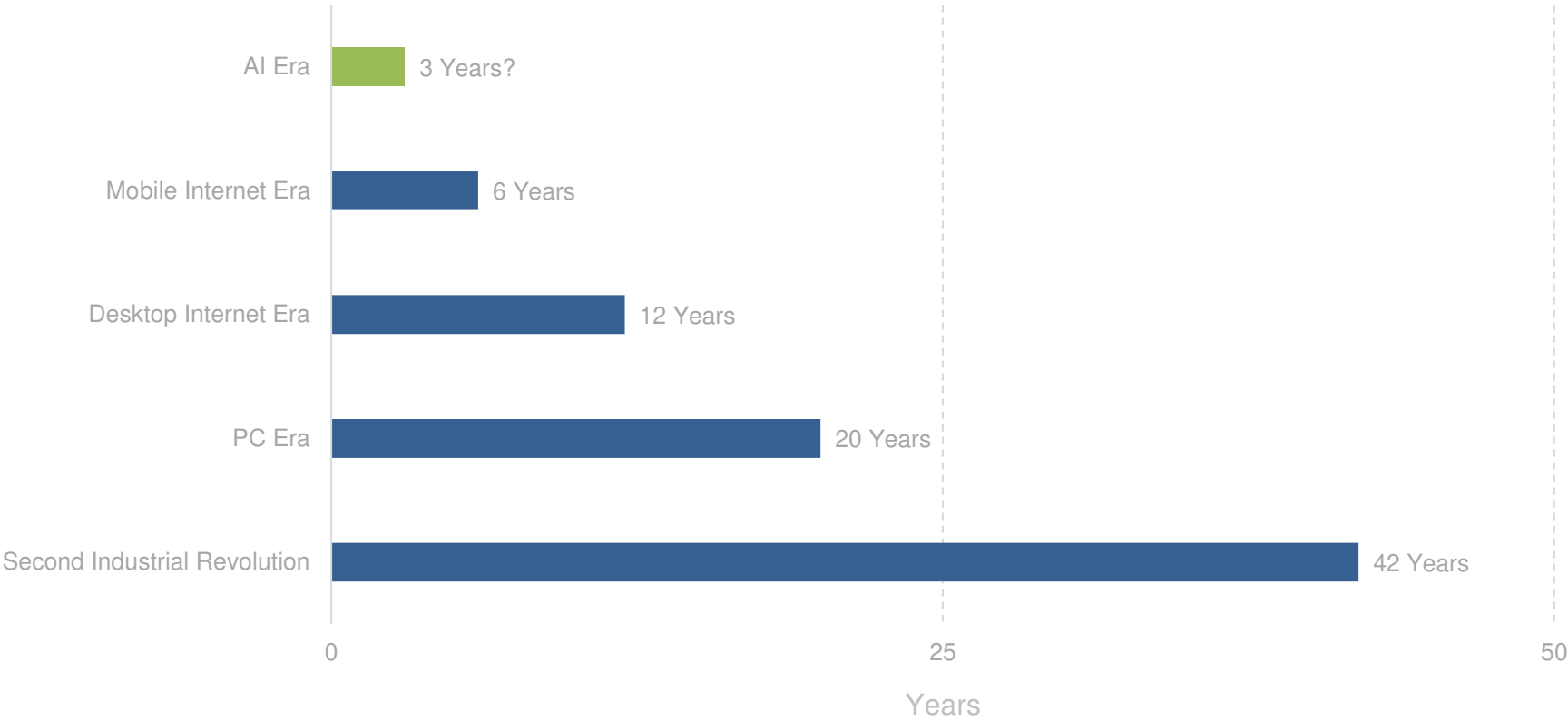
# AI 用户采用率 – 达到 50% 家庭普及率的时间 = 每个周期增速为之前的一半 ~ 时间 …AI 遵循该模式

Years to 50% Adoption of Household Technologies in USA, per Morgan Stanley



Note: 3 years for AI Era implies that the time to 50% USA Household Adoption is similarly cut in half from the previous cycle. Source: Morgan Stanley, 'Google and Meta: AI vs. Fundamental 2H Debates' (7/23), Our World in Data, other web sources per MS

Years to 50% Adoption of Household Technologies in USA, per Morgan Stanley



注：AI 时代为 3 年意味着达到 50% 美国 \*\* 家庭 \*\* 普及率的时间也比之前的周期缩短了一半。来源：摩根士丹利， “ 谷歌和 Meta：人工智能与 2H 基本面辩论 ” （ 7/23 ） ， 数据中的世界， 摩根士丹利的其他网络资源 source: Morgan Stanley, 'Google and Meta: AI vs.Fundamental 2H Debates' (7/23), Our World in Data, other web sources per MS



Technology Ecosystem AI Adoption =

Impressive

Technology Ecosystem AI Adoption =

令人印象深刻

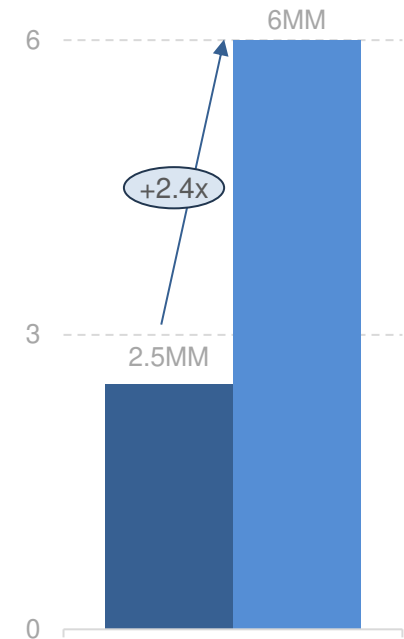
# NVIDIA AI Ecosystem Tells Over Four Years = >100% Growth in Developers / Startups / Apps

# NVIDIA AI 生态系统展示了四年内 =>100% 的开发 者 / 初创公司 / 应用增长

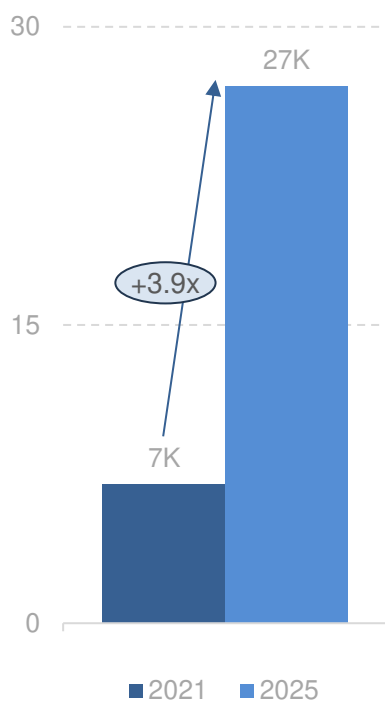
NVIDIA Computing Ecosystem – 2021-2025, per NVIDIA

NVIDIA 计算生态系统 – 2021-2025, 根据 NVIDIA

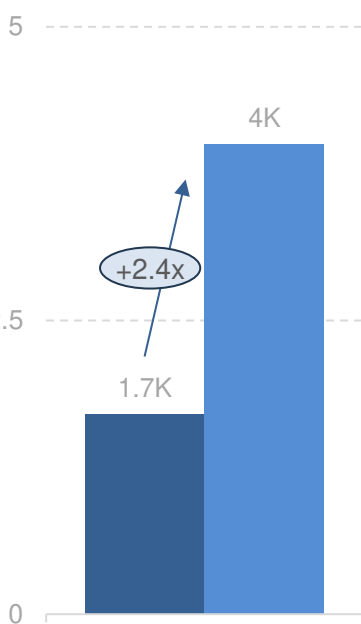
Number of Developers (MM)



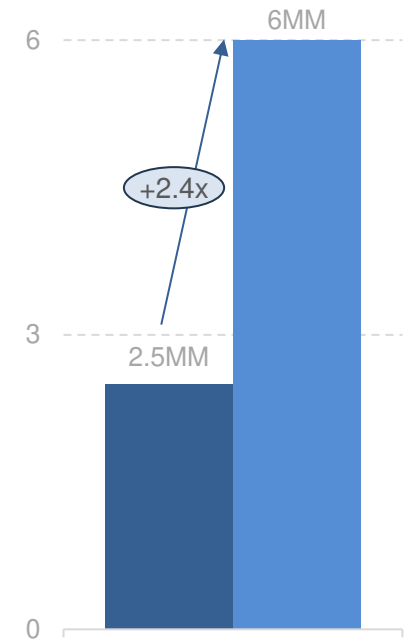
Number of AI Startups (K)



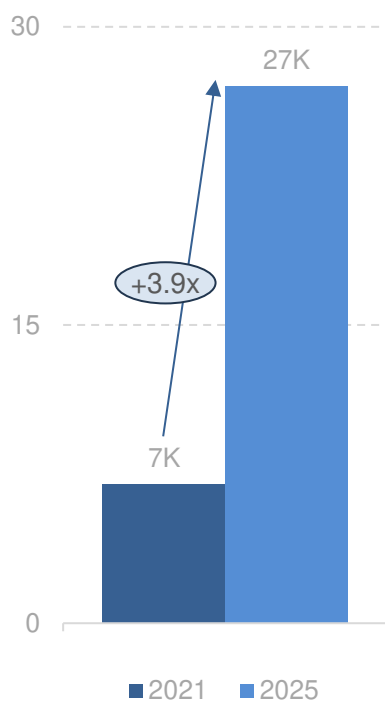
Number of Applications Using GPUs (K)



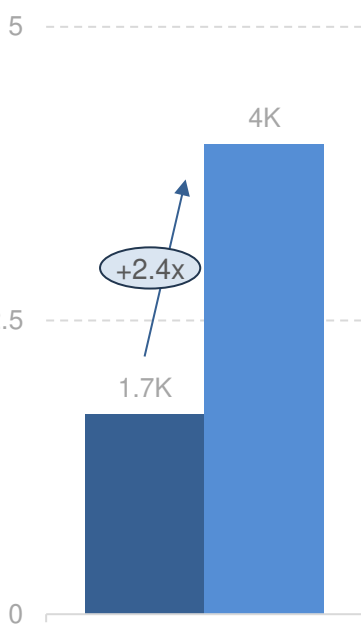
开发者数量（百万）



AI 初创公司数量（千）



使用 GPU 的应用程序数量 (K)



Note: GPU = Graphics Processing Unit. Source: NVIDIA (2021 & 2025)

注：GPU = 图形处理单元。来源：NVIDIA （2021 年和 2025 年）

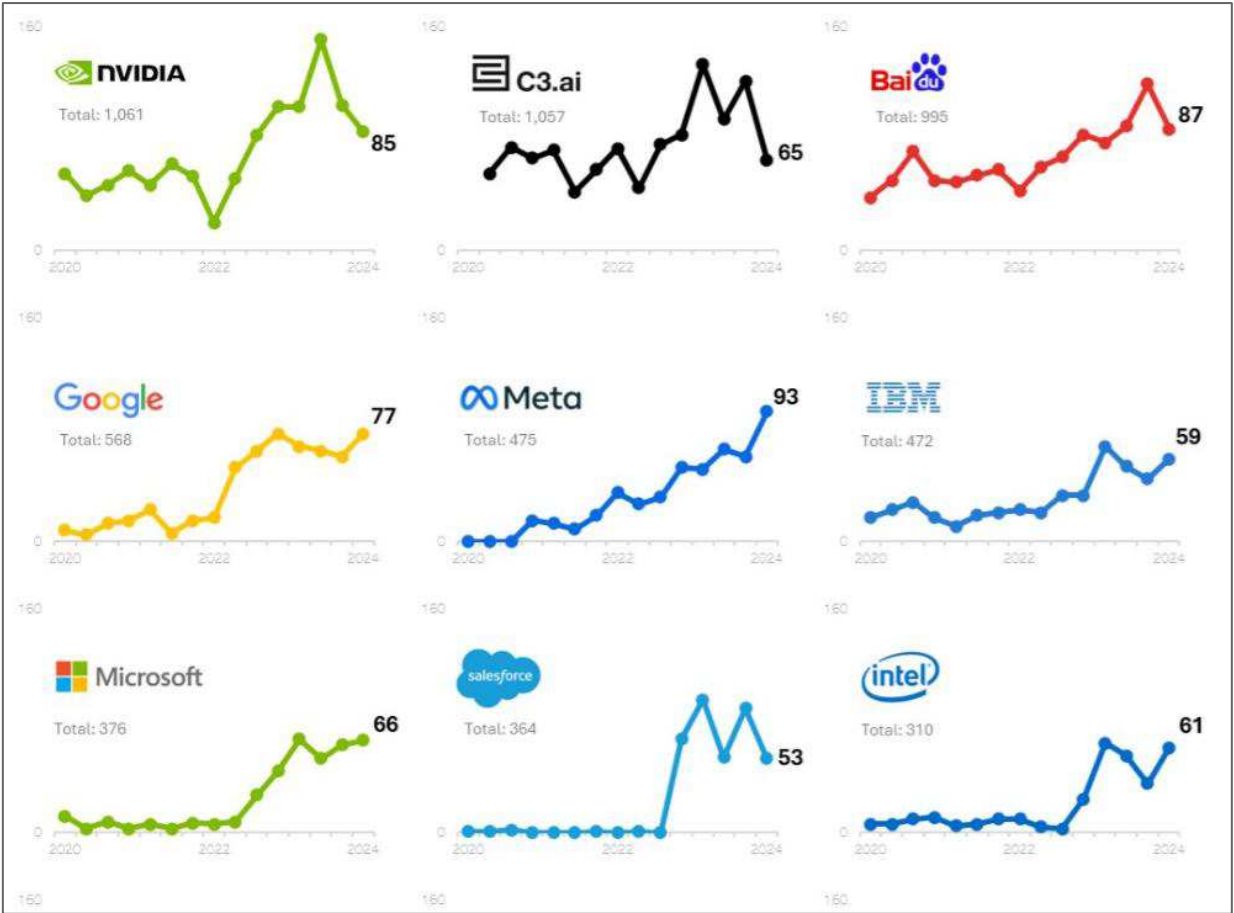
Tech Incumbent AI Adoption =  
Top Priority

科技巨头人工智能采用率 =  
首要任务

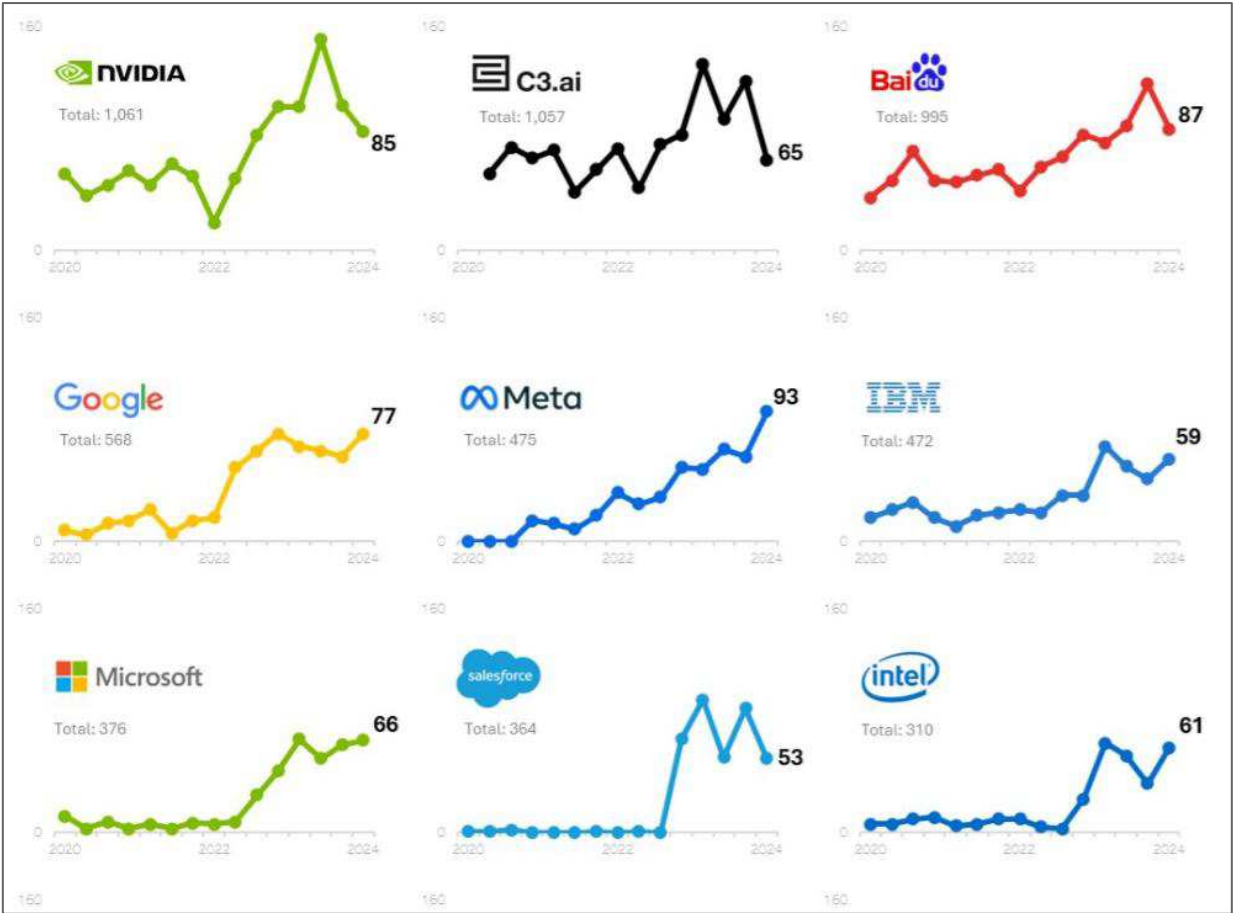
Tech Incumbent AI Focus = Talking-the-Talk...

科技巨头人工智能关注点 = 光说不练 …Talk...

Mentions of ‘AI’ in Corporate Earnings Transcripts – Q1:20-Q1:24, per Uptrends



Mentions of ‘AI’ in Corporate Earnings Transcripts – Q1:20-Q1:24, per Uptrends



Source: Uptrends, ‘Top 15 Companies Mentioning AI on Earnings Calls’ (6/24), company earnings transcripts

Source: Uptrends, ‘Top 15 Companies Mentioning AI on Earnings Calls’ (6/24), company earnings transcripts



*Generative AI is going to reinvent virtually every customer experience we know and enable altogether new ones about which we’ve only fantasized. The early AI workloads being deployed focus on productivity and cost avoidance...*

*...Increasingly, you’ll see AI change the norms in coding, search, shopping, personal assistants, primary care, cancer and drug research, biology, robotics, space, financial services, neighborhood networks – everything.*

- Amazon CEO Andy Jassy in 2024 Amazon Shareholder Letter – 4/25



生成式人工智能将重塑我们所知的几乎所有客户体验，并实现我们只能幻想的全新体验。早期部署的人工智能工作负载侧重于提高生产力和避免成本 …

… 你将越来越多地看到人工智能改变编码、搜索、购物方面的规范，个人助理、初级保健、癌症和药物研究、生物学、机器人技术、太空、金融服务、社区网络 —— 一切。

- 亚马逊首席执行官 Andy Jassy 在 2024 年亚马逊股东信中 –4/25



*The chance to improve lives and reimagine things is why Google has been investing in AI for more than a decade...*

*...We see it as the most important way we can advance our mission to organize the world's information, make it universally accessible and useful...*

*...The opportunity with AI is as big as it gets.*

- Google CEO Sundar Pichai @ Google Cloud Next 2025 – 4/25



改善生活和重新构想事物的机会是谷歌十多年来一直在投资人工智能的原因 …

… 我们认为这是我们推进组织世界信息，使其普遍可访问且有用使命的最重要方式 ……

… 人工智能带来的机遇非常巨大。

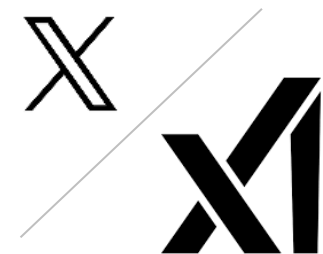
- Google CEO Sundar Pichai @ Google Cloud Next 2025 – 4/25



*There's three places where [GenAI is]...helping us:  
data creation...creating new features that were just not possible...  
efficiencies everywhere in the company...*

*...I should mention something amazing about [the new Duolingo curriculum in] chess is  
that it really started with a team of two people, neither of whom knew how to  
program...and they basically made prototypes and did the whole curriculum  
of chess by just using AI. Also, neither of them knew how to play chess.*

- Duolingo Co-Founder & CEO Luis von Ahn @ Q1:25 Earnings Call – 5/25



*AI with Grok is getting very good...it's important that AI be programmed with  
good values, especially truth-seeking values. This is, I think, essential for AI safety...*

*...Remember these words: We must have a maximally truth-seeking AI.*

- xAI Founder & CEO Elon Musk – 5/25

Note: On 3/28/25, Elon Musk announced that xAI had acquired X in an all-stock deal. The deal valued xAI at \$80B and X at \$33B (\$45B less \$12B debt). Source: Duolingo (5/1/25), DeepMind, Elon Musk (5/2/25), Fox News



*[GenAI 在 ] 以下三个方面为我们提供帮助：数据创建 …… 创  
建以前根本不可能实现的新功能 …… 提高公司各方面的效率 ……*

AI 完全回归正轨：  
DeepMind 的 AlphaGo (2014) 开  
始于人类训练机器 ……Duolingo  
Chess 现在让机器训练人类 ……

… 我应该提一下关于 [Duolingo 新的国际象棋课程 ] 的一些令人惊奇的事情，它最初是由一个两人团队开始的，他们两人都不知道如何编程 …… 他们基本上制作了原型，并完全通过使用人工智能完成了整个国际象棋课程。而且，他们两人都不知道如何下国际象棋。

- Duolingo 联合创始人兼首席执行官 Luis von Ahn @ 25 年第一季度财报电话会议 – 5/25



使用 Grok 的 AI 变得非常出色 …… 重要的是，AI 应该被编程为具有良好的价值观，尤其是追求真理的价值观。我认为，这对于 AI 安全至关重要 ……

… 记住这些话：我们必须拥有一个最大限度追求真理的 AI。

- xAI 创始人兼 CEO Elon Musk –5/25

注意：在 25 年 3 月 28 日，Elon Musk 宣布 xAI 已通过全股票交易收购 X。该交易对 xAI 的估值为 800 亿美元，对 X 的估值为 330 亿美元（450 亿美元减去 120 亿美元债务）。来源：Duolingo (25 年 5 月 1 日)、DeepMind、Elon Musk (25 年 5 月 2 日)、Fox News



*We view AI as a human acceleration tool that will allow individuals to do more... I believe long term, we will see people coupled with... the AI they use as the overall output of that person.*

**- Roblox Co-Founder, President, CEO & Chair of Board David Baszucki @ Q1:25 Earnings Call – 5/25**



我们认为人工智能是一种人类加速工具， 它将使个人能够做更多的事情我相信从长远来 ... 看， 我们会看到人们与 …… 相结合 *the AI they use as the overall output of that person.*

**- Roblox 联合创始人、总裁、首席执行官兼董事会主席 David Baszucki @ Q1:25 财报电话会议 – 5/25**



*I promise you, in ten years' time, you will look back and you will realize that AI has now integrated into everything. And in fact, we need AI everywhere.*

*And every region, every industry, every country, every company, all needs AI. AI [is] now part of infrastructure. And this infrastructure, just like the internet, just like electricity, needs factories....*

*...And these AI data centers, if you will, are improperly described. They are, in fact, AI factories. You apply energy to it, and it produces something incredibly valuable.*

**- NVIDIA Co-Founder & CEO Jensen Huang @ COMPUTEX 2025 – 5/25**



我向你保证，在十年后，当你回首往事时，你会意识到人工智能现在已经融入到一切事物中。事实上，我们需要在所有地方都使用人工智能。

每个地区、每个行业、每个国家、每家公司都需要人工智能。人工智能 [is] 现在是基础设施的一部分。这种基础设施，就像互联网，就像电力一样，需要工厂 ……。

…… 这些人工智能数据中心，如果你愿意这么称呼它们的话，它们的描述是不准确的。事实上，它们是人工智能工厂。你向其中注入能量，它就会产生非常有价值的东西。

**- NVIDIA 联合创始人兼首席执行官黄仁勋 @ COMPUTEX 2025 – 5/25**

Source: Roblox (5/1/25), NVIDIA (5/18/25)

来源: Roblox (5/1/25), NVIDIA (5/18/25)

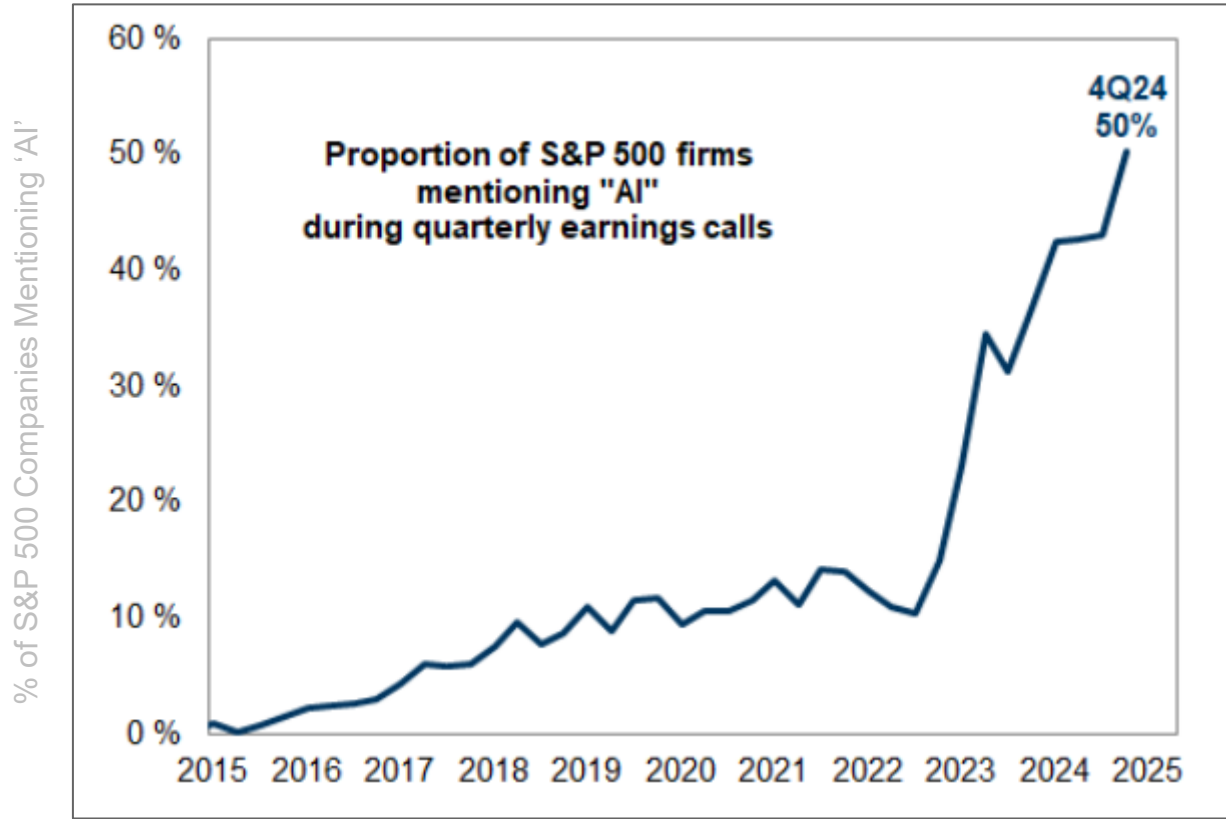
‘Traditional’ Enterprise AI Adoption =  
Rising Priority

“传统” 企业 AI 采用 =  
日益重要的优先级



Enterprise AI Focus – S&P 500 Companies = 50% & Rising Talking-the-Talk

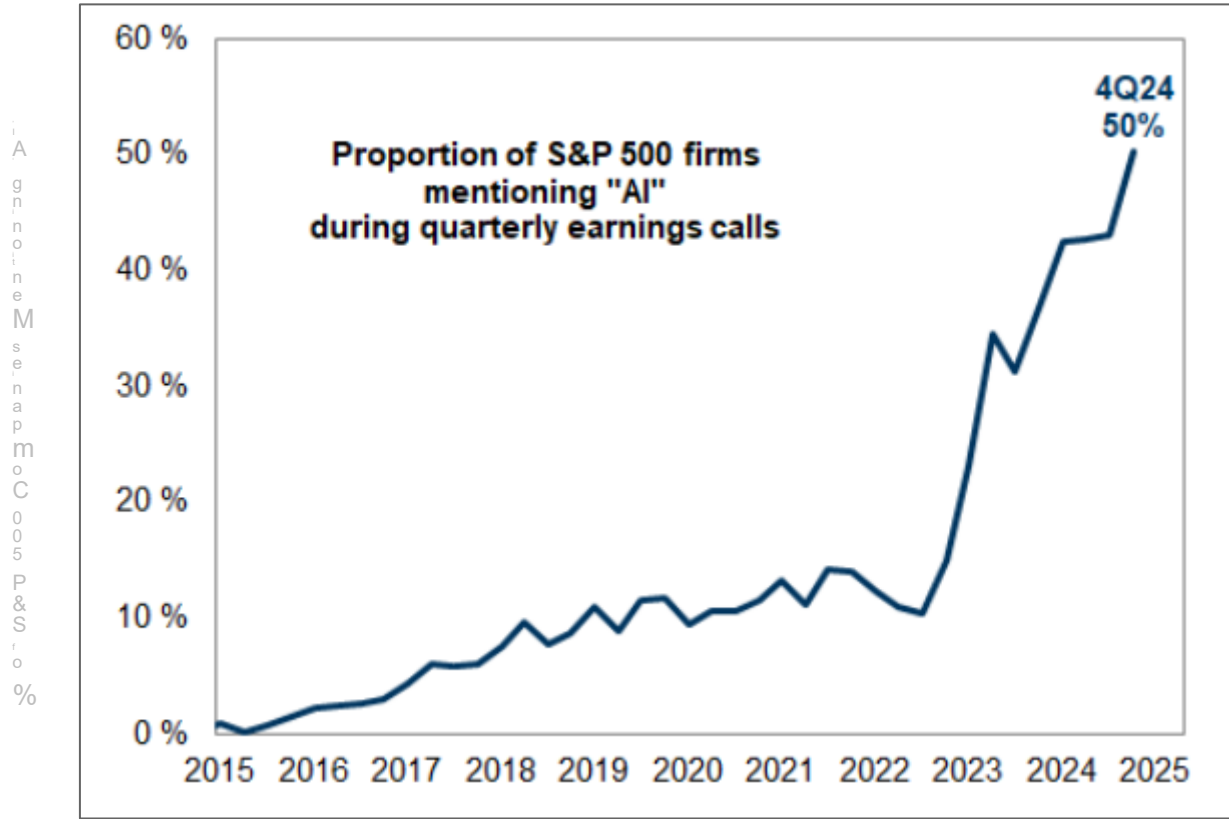
Quarterly Earnings Call Mentions of ‘AI’ – S&P 500 Companies (2015-2025), per Goldman Sachs Research



Source: Goldman Sachs Global Investment Research, 'S&P Beige Book: 3 themes from 4Q 2024 conference calls: Tariffs, a stronger US dollar, and AI' (2/25)

Enterprise AI Focus – S&P 500 Companies = 50% & Rising Talking-the-Talk

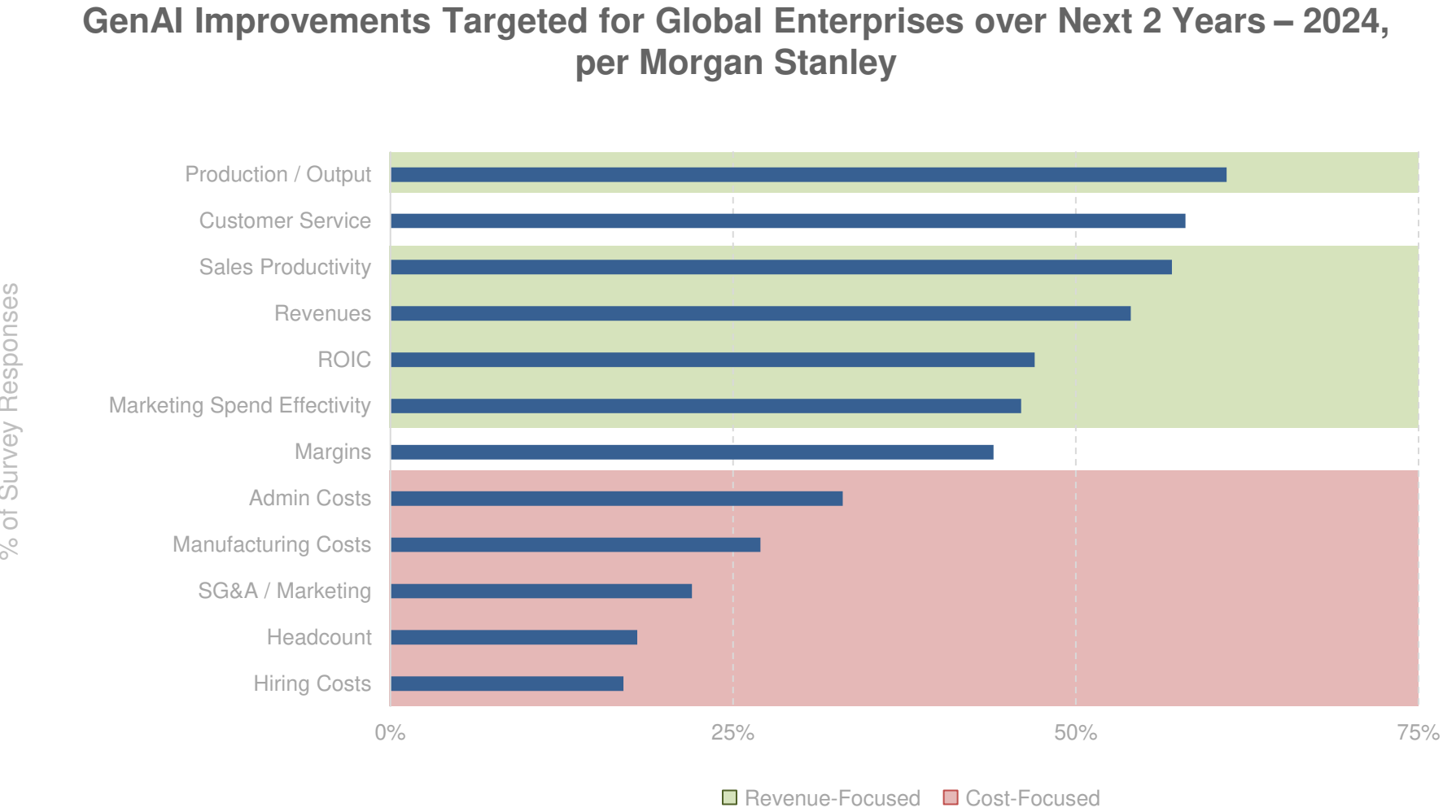
Quarterly Earnings Call Mentions of ‘AI’ – S&P 500 Companies (2015-2025), per Goldman Sachs Research



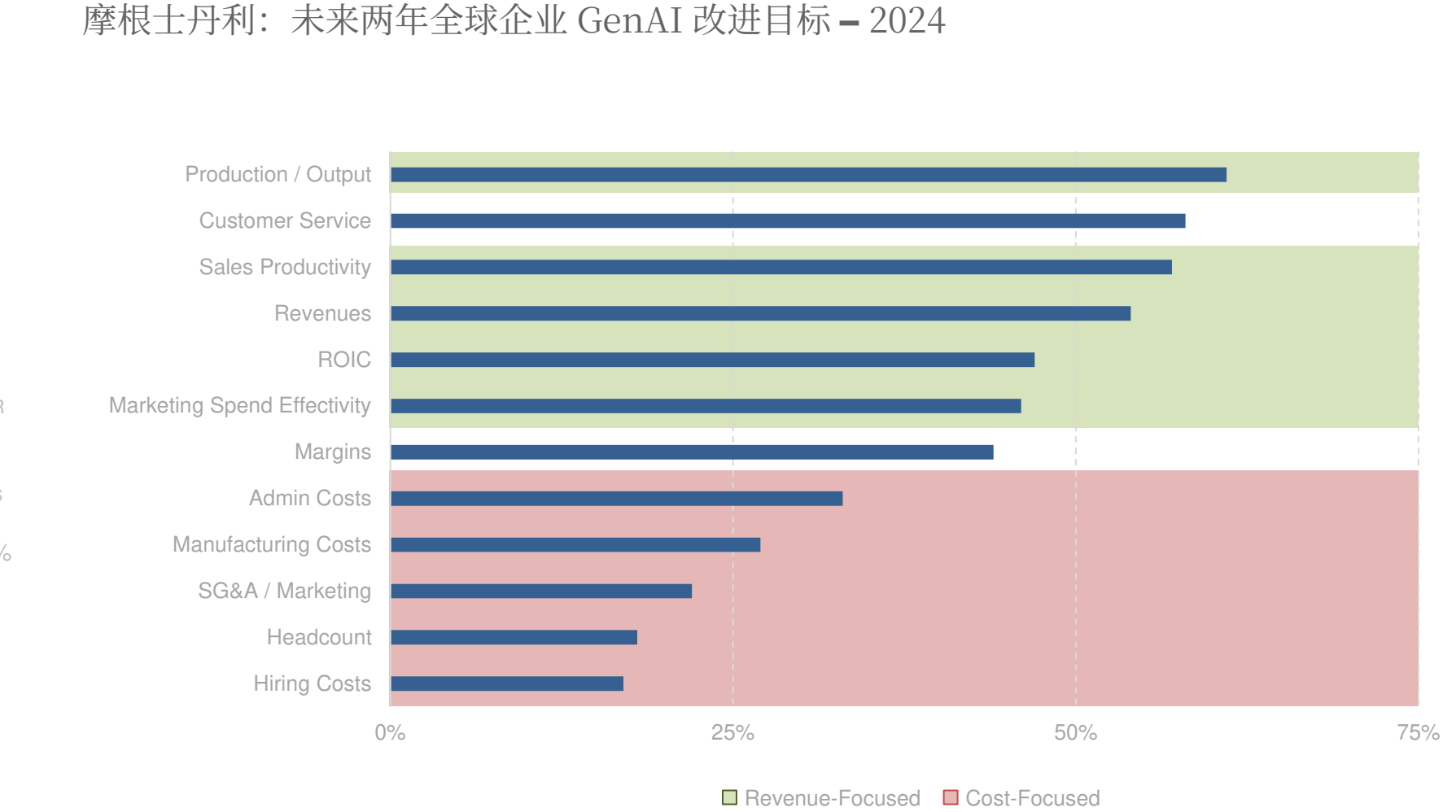
Source: Goldman Sachs Global Investment Research, 'S&P Beige Book: 3 themes from 4Q 2024 conference calls: Tariffs, a stronger US dollar, and AI' (2/25)

Enterprise AI Focus – Global Enterprises = Growth & Revenue...Not Cost Reduction

Enterprise AI Focus – Global Enterprises = Growth & Revenue...Not Cost Reduction



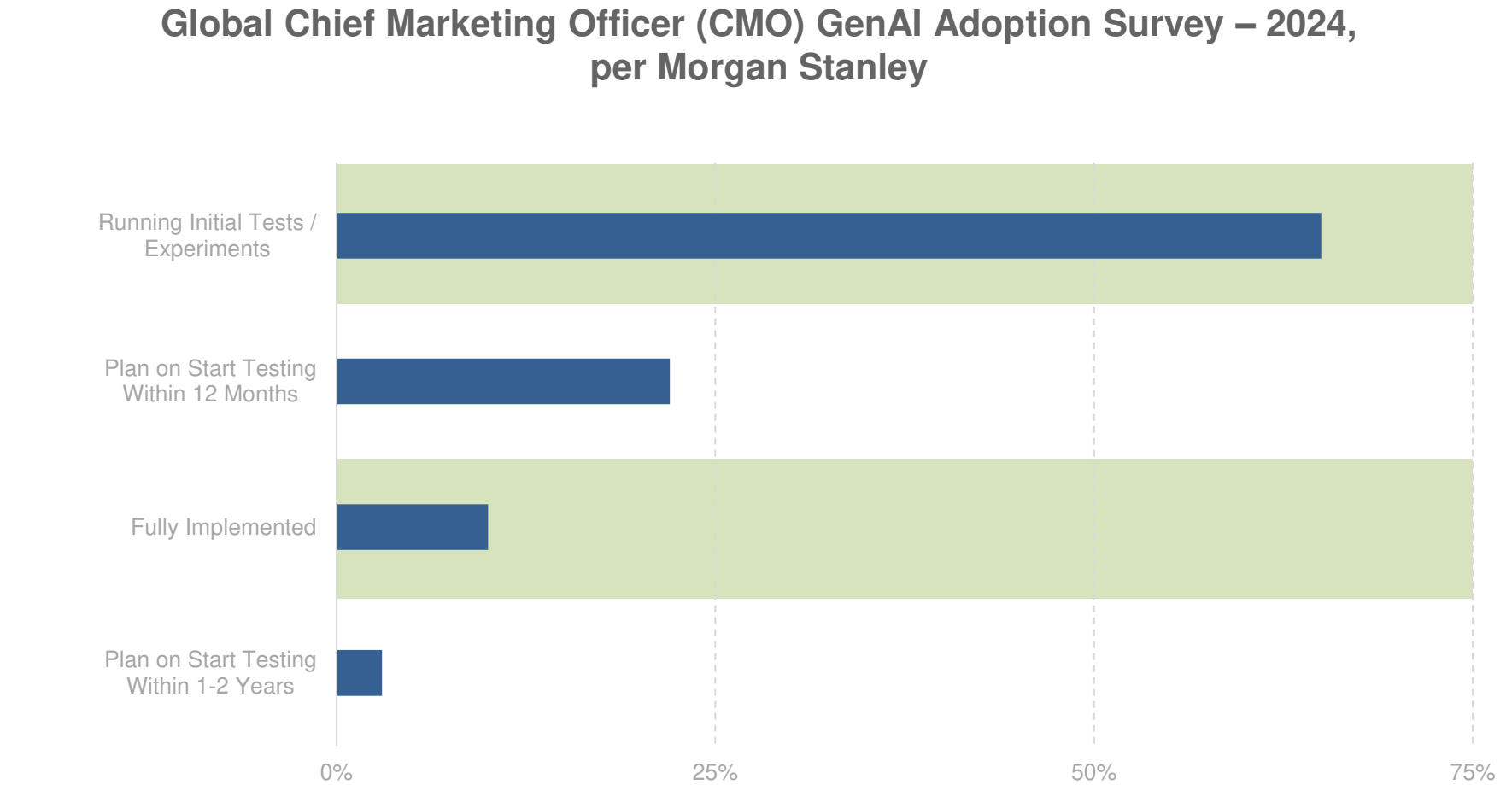
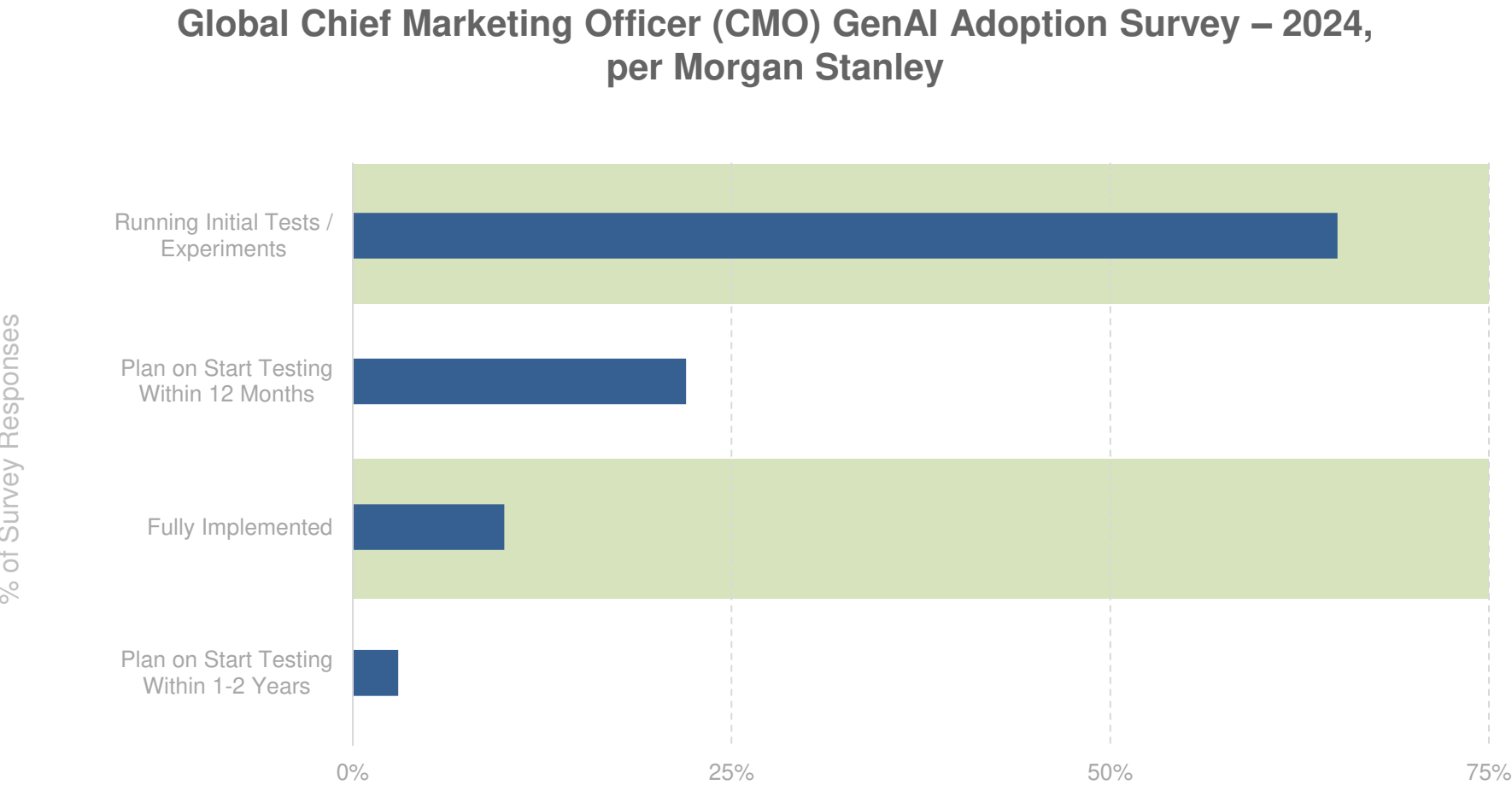
Note: Survey conducted 5/24, N=427. US-based companies = 43%, Japan 15%, UK 14%, France 14%, Germany 14%. Industry mix: 18% Technology, 18% Financial Services, 17% Healthcare, 17% Manufacturing, 15% Industrials, 15% Consumer,. Revenue mix: 13% \$500MM-\$750MM, 25% \$751MM-\$1B, 36% \$1B-\$5B, 10% \$5B-\$10B, 8% \$10B-\$15B, 3% \$15B-\$20B, 5% \$20B+. 'Revenue-Focused' and 'Cost-Focused' categorizations per BOND, not Morgan Stanley. Source: AlphaWise, Morgan Stanley, 'Quantifying the AI Opportunity' (12/24)



注意：调查于 5/24 进行，N=427。美国公司 = 43%，日本 15%，英国 14%，法国 14%，德国 14%。行业构成：18% 技术，18% 金融服务，17% 医疗保健，17% 制造业，15% 工业，15% 消费。收入构成：13% 5 亿美元 - 7.5 亿美元，25% 7.51 亿美元 - 10 亿美元，36% 10 亿美元 - 50 亿美元，10% 50 亿美元 - 100 亿美元，8% 100 亿美元 - 150 亿美元，3% 150 亿美元 - 200 亿美元，5% 200 亿美元 +。“收入 - 导向”和“成本 - 导向”的分类来自 BOND，而非摩根士丹利。来源：AlphaWise，摩根士丹利，《量化人工智能机遇》(12/24)

Enterprise AI Focus – Global CMOs = 75% Using / Testing AI Tools

企业 AI 焦点 – 全球 CMO =75% 正在使用 / 测试 AI 工具



Note: Survey question asked about the extent to which marketing executives worldwide are using generative AI for marketing activities. Survey conducted 7/24, N = 300 marketing executives at companies with 500+ employees worldwide. Survey geos: Australia, Belgium, Brazil, Canada, China, Denmark, Finland, France, Germany, Ireland, Italy, Japan, Luxembourg, Mexico, Netherlands, Norway, Poland, Saudi Arabia, Spain, Sweden, UAE, UK, & USA. Source: eMarketer, Morgan Stanley, 'Quantifying the AI Opportunity' (12/24)

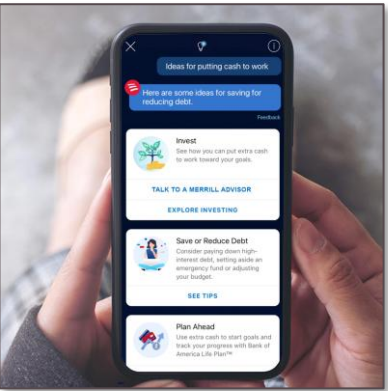
注意：调查问题询问了全球营销主管在多大程度上使用生成式 AI 进行营销活动。调查于 7/24 进行，N = 300 营销主管来自在全球拥有 500+ 名员工的公司。调查地区：澳大利亚、比利时、巴西、加拿大、中国、丹麦、芬兰、法国、德国、爱尔兰、意大利、日本、卢森堡、墨西哥、荷兰、挪威、波兰、沙特阿拉伯、西班牙、瑞典、阿联酋、英国和美国。来源：eMarketer、摩根士丹利，《量化 AI 机遇》(12/24)》

Enterprise AI Adoption = Rising Priority...  
Bank of America – Erica Virtual Assistant (6/18)

企业人工智能应用 = 日益受到重视 ··· 美国银行 –  
Erica 虚拟助手（ 6/18 ）

Bank of America Erica Virtual Assistant – 6/18-2/25,  
per Bank of America

美国银行 Erica 虚拟助手 – 6/18-2/25，数据来源：美国银行

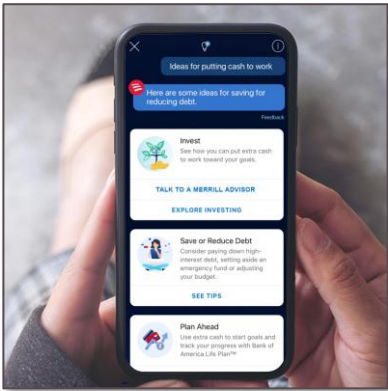
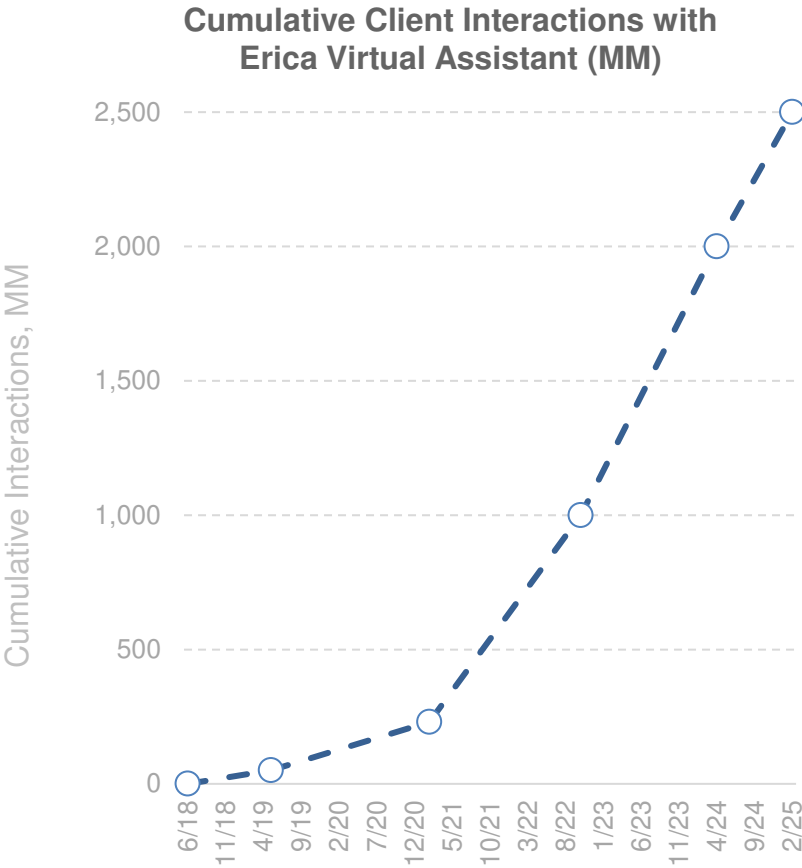


Note: Erica is a conversational AI built into Bank of America’s mobile app that helps customers manage their finances by providing real-time insights, transaction search, bill reminders, and budgeting assistance. It has handled billions of interactions and serves as a 24/7 digital financial concierge for over 40 million clients.

Erica acts as both a personal concierge and mission control for our clients.

Our data science team has made more than 50,000 updates to Erica’s performance since launch – adjusting, expanding and fine-tuning natural language understanding capabilities, ensuring answers and insights remain timely and relevant. 2 billion client interactions is a compelling milestone though this is only the beginning for Erica.

- Head of Digital at Bank of America Nikki Katz, 4/24

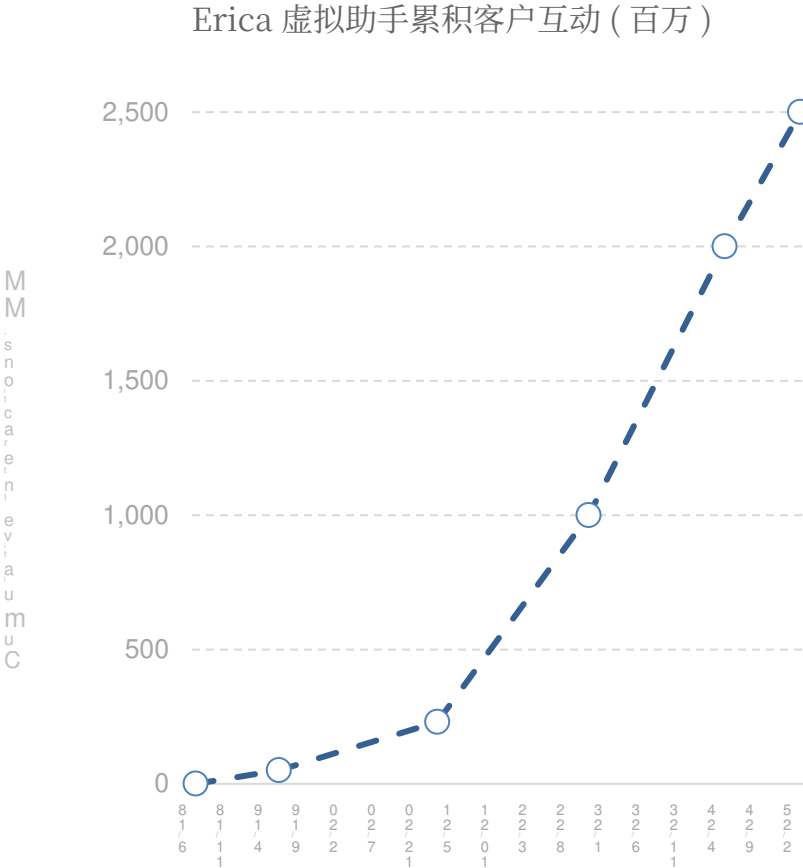


注意：Erica 是一种会话式人工智能，内置于美国银行的移动应用程序中，通过提供实时见解、交易搜索、账单提醒和预算帮助来帮助客户管理财务。它已处理数十亿次互动，并为超过 4000 万客户提供 24/7 全天候的数字金融礼宾服务。

Erica 既可以作为我们客户的私人礼宾，也可以作为任务控制中心。

自发布以来，我们的数据科学团队对 Erica 的性能进行了超过 50,000 次的更新 – 调整、扩展和微调自然语言理解能力，确保答案和见解保持及时和相关性。 20 亿次客户互动是一个令人信服的里程碑，但这仅仅是 Erica 的开始。

- 美国银行数字主管 Nikki Katz, 4/24



Note: We assume a start at zero users from Erica’s launch in 6/18. Pilot users excluded. Source: Bank of America (2/21, 4/24, 2/25)

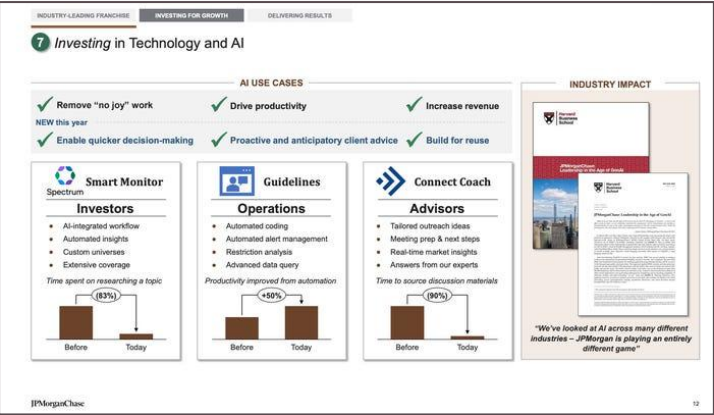
Note: We 假设 Erica 于 6/18 发布时，用户数量从零开始。不包括试点用户。来源：美国银行（ 2/21 、 4/24 、 2/25 ）

Enterprise AI Adoption = Rising Priority...  
JP Morgan – End-to-End AI Modernization (2020)

Enterprise AI Adoption = Rising Priority...  
JP Morgan – End-to-End AI Modernization (2020)

JP Morgan End-to-End AI Modernization – 2023-2025E,  
per JP Morgan

JP Morgan End-to-End AI Modernization – 2023-2025E,  
根据摩根大通



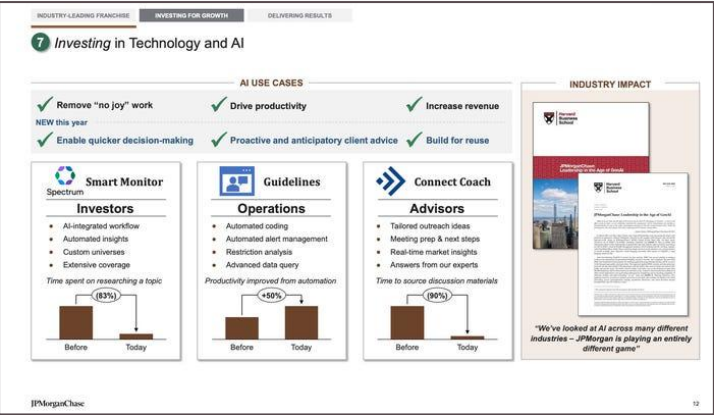
We have high hopes for the efficiency gains we might get [from AI]...

...Certain key subsets of the users tell us they are gaining several hours a week of productivity, and almost by definition, the time savings is coming from less valuable tasks...

...We were early movers in AI. But we're still in the early stages of the journey.

- JP Morgan CFO Jeremy Barnum, 5/25

Note: Superscript '2', per JP Morgan, indicates 'Value is described as benefit in revenue, lower expense, or avoidance of cost – majority is measured as the lift relative to prior analytical techniques with the remainder relative to a random baseline or holdout control.' We indicate 2020 as the start year for JP Morgan's AI Modernization (2020 Letter to Shareholders: 'We already extensively use AI, quite successfully, in fraud and risk, marketing, prospecting, idea generation, operations, trading and in other areas—to great effect, but we are still at the beginning of this journey'). Source: JP Morgan Investor Day (5/25)



我们对 [ 人工智能 ]…… 可能带来的效率提升寄予厚望

…… 某些关键用户子集告诉我们，他们每周的生产力提高了几个小时，而且几乎可以肯定的是，节省的时间来自价值较低的任务 ……

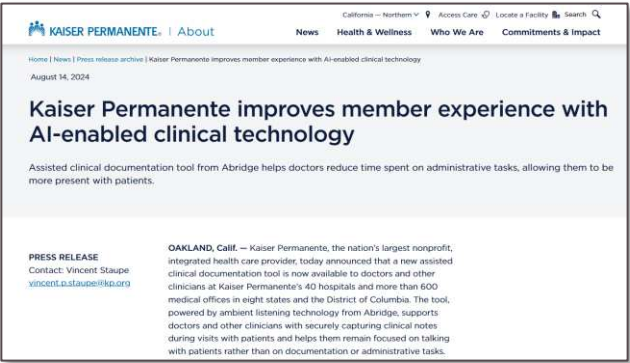
…… 我们是人工智能领域的先行者。但我们仍处于这段旅程的早期阶段。

- JP Morgan CFO Jeremy Barnum, 5/25

注意：根据摩根大通的说法，上标“2”表示“价值被描述为收入、较低费用或避免成本方面的收益 – 大多数是相对于先前分析技术的提升来衡量的，其余部分是相对于随机基线或保留控制来衡量的。我们将 2020 年定为摩根大通人工智能现代化的开始年份（2020 年致股东的信：‘我们已经在欺诈和风险、营销、勘探、创意生成、运营、交易和其他领域广泛且成功地使用人工智能 – 效果很好，但我们仍处于这段旅程的开始阶段’）。资料来源：摩根大通投资者日（5/25）

# Enterprise AI Adoption = Rising Priority... Kaiser Permanente – Multimodal Ambient AI Scribe (10/23)

## Kaiser Permanente Ambient AI Scribe – 10/23-12/24, per New England Journal of Medicine



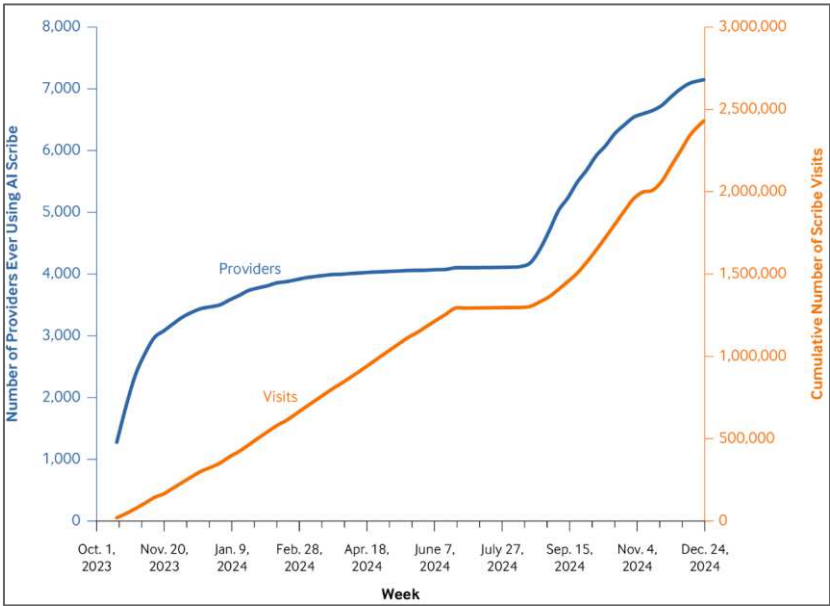
Ambient artificial intelligence (AI) scribes, which use machine learning applied to conversations to facilitate scribe-like capabilities in real time, [have] great potential to reduce documentation burden, enhance physician-patient encounters, and augment clinicians’ capabilities.

The technology leverages a smartphone microphone to transcribe encounters as they occur but does not retain audio recordings. To address the urgent and growing burden of data entry, in October 2023, The Permanente Medical Group (TPMG) enabled ambient AI technology for 10,000 physicians and staff to augment their clinical capabilities across diverse settings and specialties.

- New England Journal of Medicine Catalyst Research Report, 2/24

Source: Tierney, Aaron A. et al., ‘Ambient Artificial Intelligence Scribes to Alleviate the Burden of Clinical Documentation’ (3/24) & Tierney, Aaron A. et al., ‘Ambient Artificial Intelligence Scribes: Learnings after 1 Year and over 2.5 Million Uses’ (3/25) via Nestor Maslej et al., ‘The AI Index 2025 Annual Report,’ AI Index Steering Committee, Stanford HAI (4/25)

## Unique Kaiser Permanente Physicians Ever Using AI Scribe & Cumulative Number of Scribe Visits



# 企业 AI 采用 = 日益重要的优先事项 ……Kaiser Permanente – 多模态环境 AI 记录器 (10/23)

## Kaiser Permanente Ambient AI Scribe – 10/23-12/24, per New England Journal of Medicine



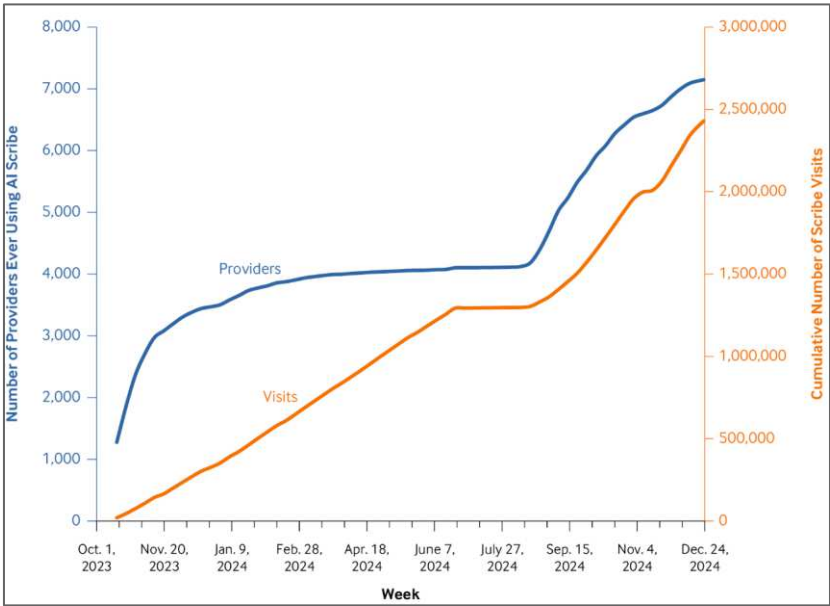
环境人工智能 (AI) 记录器，它使用应用于对话的机器学习来实时促进类似记录器的功能，[有] 巨大的潜力来减轻文档负担、增强医患互动并增强临床医生的能力。

该技术利用智能手机麦克风来转录发生的互动，但不保留录音。为了解决紧迫且不断增长的数据录入负担，2023 年 10 月，Permanente 医疗集团 (TPMG) 为 10,000 名医生和员工启用了环境 AI 技术，以增强他们在各种环境和专业中的临床能力。

- 《新英格兰医学杂志催化剂》研究报告，2/24

来源：Tierney, Aaron A. 等人，《利用环境人工智能记录器减轻临床文档负担》(3/24) & Tierney, Aaron A. 等人，《环境人工智能记录器：一年后和超过 250 万次使用后的经验》(3/25)，来自 Nestor Maslej 等人，《2025 年人工智能指数年度报告》，人工智能指数指导委员会，斯坦福 HAI (4/25)

## 使用 AI Scribe 的 Kaiser Permanente 独特医生数量 & Scribe 访问累计次数

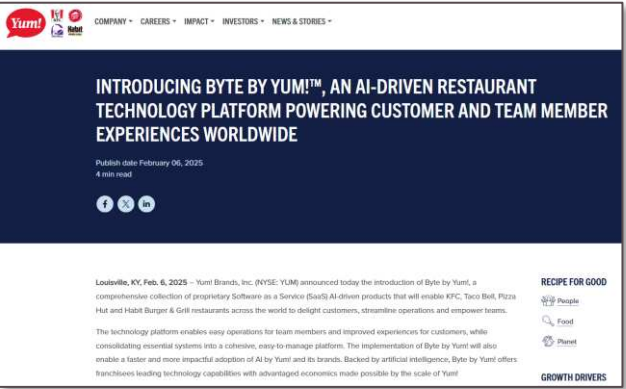




Enterprise AI Adoption = Rising Priority...  
Yum! Brands – Byte by Yum! (2/25)

企业 AI 采用 = 日益重要的优先事项 ...Yum!  
Brands – Byte by Yum! (2/25)

Yum! Brands Byte by Yum! – 2/24-2/25, per Yum! Brands

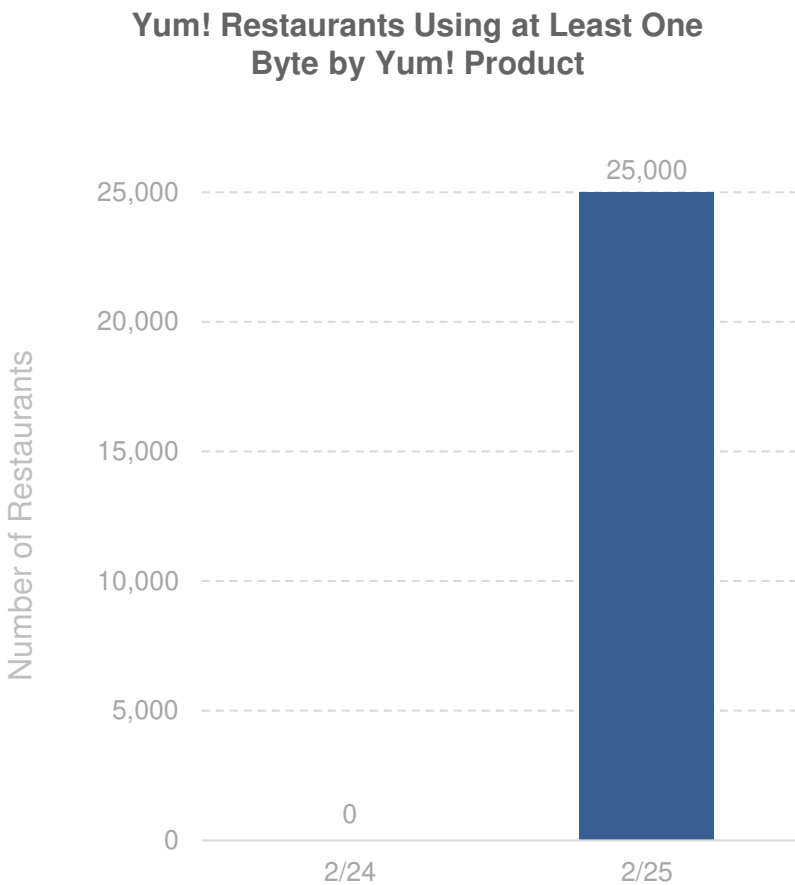


Byte is Yum! Brands' AI-powered restaurant management platform designed to optimize store operations by automating repetitive tasks like inventory tracking, scheduling, and food preparation alerts. It leverages machine learning to improve decision-making at the restaurant level, enhancing efficiency, reducing waste, and supporting staff productivity.

*Backed by artificial intelligence, Byte by Yum! offers franchisees leading technology capabilities with advantaged economics made possible by the scale of Yum!.*

*The Byte by Yum! platform includes online and mobile app ordering, point of sale, kitchen and delivery optimization, menu management, inventory and labor management, and team member tools.*

**- Yum! Press Release, 2/25**



Note: Yum! Brands names include KFC, Taco Bell, Pizza Hut, & The Habit. Byte by Yum! was officially launched in 2/25. While underlying technologies were previously in-use at restaurants in Yum!’s portfolio, the Byte by Yum! product suite had not yet officially been launched; hence, we illustratively show zero users in 2/24. Source: Yum!, ‘Introducing Byte by Yum!™, an AI-driven restaurant technology platform powering customer and team member experiences worldwide’ (2/25)

Yum! Brands Byte by Yum! – 2/24-2/25, 数据来源：Yum! Brands



Byte 是 Yum! Brands 的人工智能餐厅管理平台，旨在通过自动化重复性任务（如库存跟踪、排班和食品准备警报）来优化门店运营。它利用机器学习来改善餐厅层面的决策，从而提高效率、减少浪费并支持员工生产力。

在人工智能的支持下，Byte by Yum! 为加盟商提供领先的技术能力，并凭借 Yum! 的规模实现有利的经济效益。

Byte by Yum! 平台包括在线和移动应用订购、销售点、厨房和配送优化、菜单管理、库存和劳动力管理以及团队成员工具。

**- Yum! 新闻稿，2/25**

注意：Yum! Brands 的品牌包括肯德基、塔可钟、必胜客和 The Habit。Byte by Yum! 于 2 月 25 日正式推出。虽然底层技术之前已在 Yum! 旗下的餐厅中使用，但 Byte by Yum! 产品套件尚未正式推出；因此，我们在 2/24 中示意性地显示零用户。来源：Yum!, “Introducing Byte byYum!, 一个由 AI 驱动的餐厅技术平台，为全球客户和团队成员提供支持 ” （2/25）

Education / Government / Research AI Adoption =

Rising Priority

Education / Government / Research AI Adoption =

日益重要的优先事  
项




Education & Government = Increasingly Announcing AI Integrations

教育和政府部门 = 越来越多地宣布人工智能集成

Arizona State University's 'AI Acceleration' – 8/23

New team of technologists creating artificial intelligence (AI) tools



ASU Enterprise Technology

ASU unveils first-of-its-kind organization to advance AI across the enterprise

August 2, 2023

Recognized as the nation's most innovative university by U.S. News and World Report for eight years and counting, ASU is at the forefront of cutting-edge technology. In continuing to chart the pathway of tech and innovation, ASU's Enterprise Technology announced today the launch of AI Acceleration, a new team of technologists dedicated to creating the next generation of artificial intelligence (AI) tools.

The group is chartered with providing positive outcomes for students, paving new pathways for groundbreaking research, and developing the technical infrastructure for more efficient business operations.

AI Acceleration is part of a larger initiative being led by ASU's Enterprise Technology to drive strategy across the university that empowers all students, faculty and staff to leverage the advantages of AI to enhance their daily work.

Under the leadership of Executive Director Elizabeth Rieffley and AI Technology Director Zuhair Zaidi, AI Acceleration brings together a team of 15 people working as machine learning operations (ML Ops) engineers, data scientists and AI development engineers to drive the development of AI platforms and products.

"Our goal is to develop AI platforms and products that effectively cater to the needs of students and society," says Rieffley. "This encompasses technologies that enhance learning, knowledge, and overall progress for everyone, along with tools that improve the accessibility, affordability and personalization of higher education."

Work under the AI Acceleration team is already underway. For example, a new platform is being developed that provides next-generation AI technology to teams across the university. The platform is designed to accomplish various tasks, such as training and deploying large and discreet language models, natural language processing (NLP), predictive modeling and more – regardless of their knowledge of software development. By leveraging these models, the university can make informed decisions that enhance retention, enrollment, operations, and student experience.

Oxford Partnership – 3/25

5-Year Partnership on Research & AI Literacy



UNIVERSITY OF OXFORD

NEWS & EVENTS

Oxford and OpenAI launch collaboration to advance research and education

Home / News / Oxford and OpenAI launch collaboration to advance research and education

Published: 4 MAR 2023


Share this

ARTIFICIAL INTELLIGENCE EDUCATION RESEARCH UNIVERSITY

The University of Oxford has announced plans to expand its artificial intelligence (AI) offering and capabilities with OpenAI.

NextGenAI – 3/25

\$50MM consortium with 15 research universities (MIT, Harvard, Caltech, etc.)



The University of Michigan RECORD

News for faculty, staff and retirees

March 4, 2025

Share on: X f

U-M, OpenAI launch partnership to expand AI research

By Don Jordan  
Office of the Vice President for Research

Topics: Campus News, Research

The University of Michigan has established a partnership with OpenAI that will bring additional artificial intelligence resources, research funding and computing power to campus.

The collaboration also will include joint research projects between the university and OpenAI, focusing on AI applications that broadly benefit society.

亚利桑那州立大学的 “人工智能加速” – 8/23

创建人工智能 (AI) 工具的全新技术专家团队



ASU Enterprise Technology

ASU unveils first-of-its-kind organization to advance AI across the enterprise

August 2, 2023

Recognized as the nation's most innovative university by U.S. News and World Report for eight years and counting, ASU is at the forefront of cutting-edge technology. In continuing to chart the pathway of tech and innovation, ASU's Enterprise Technology announced today the launch of AI Acceleration, a new team of technologists dedicated to creating the next generation of artificial intelligence (AI) tools.

The group is chartered with providing positive outcomes for students, paving new pathways for groundbreaking research, and developing the technical infrastructure for more efficient business operations.

AI Acceleration is part of a larger initiative being led by ASU's Enterprise Technology to drive strategy across the university that empowers all students, faculty and staff to leverage the advantages of AI to enhance their daily work.

Under the leadership of Executive Director Elizabeth Rieffley and AI Technology Director Zuhair Zaidi, AI Acceleration brings together a team of 15 people working as machine learning operations (ML Ops) engineers, data scientists and AI development engineers to drive the development of AI platforms and products.

"Our goal is to develop AI platforms and products that effectively cater to the needs of students and society," says Rieffley. "This encompasses technologies that enhance learning, knowledge, and overall progress for everyone, along with tools that improve the accessibility, affordability and personalization of higher education."

Work under the AI Acceleration team is already underway. For example, a new platform is being developed that provides next-generation AI technology to teams across the university. The platform is designed to accomplish various tasks, such as training and deploying large and discreet language models, natural language processing (NLP), predictive modeling and more – regardless of their knowledge of software development. By leveraging these models, the university can make informed decisions that enhance retention, enrollment, operations, and student experience.

牛津合作关系 – 3/25

5-Year Partnership on Research & AI Literacy



UNIVERSITY OF OXFORD

NEWS & EVENTS

Oxford and OpenAI launch collaboration to advance research and education

Home / News / Oxford and OpenAI launch collaboration to advance research and education

Published: 4 MAR 2023


Share this

ARTIFICIAL INTELLIGENCE EDUCATION RESEARCH UNIVERSITY

The University of Oxford has announced plans to expand its artificial intelligence (AI) offering and capabilities with OpenAI.

下一代人工智能 – 3/25

拥有 15 所研究型大学（麻省理工学院、哈佛大学、加州理工学院等）的 5000 万美元联盟



The University of Michigan RECORD

News for faculty, staff and retirees

March 4, 2025

Share on: X f

U-M, OpenAI launch partnership to expand AI research

By Don Jordan  
Office of the Vice President for Research

Topics: Campus News, Research

The University of Michigan has established a partnership with OpenAI that will bring additional artificial intelligence resources, research funding and computing power to campus.

The collaboration also will include joint research projects between the university and OpenAI, focusing on AI applications that broadly benefit society.

ChatGPT Gov – 1/25

ChatGPT tailored for USA federal agencies



launch

Solutions Sectors Studios Insights Partners About Us

OpenAI Introduces ChatGPT Gov: AI-Powered Innovation for Government Agencies

USA National Laboratories – 1/25

Partnering on Nuclear, Cybersecurity, & Scientific Breakthroughs



KUUM

89.9 FM Live From The University of New Mexico

OpenAI touts new government partnership and support for A.I. infrastructure

ChatGPT Gov – 1/25

ChatGPT tailored for USA federal agencies



launch

Solutions Sectors Studios Insights Partners About Us

OpenAI Introduces ChatGPT Gov: AI-Powered Innovation for Government Agencies

美国国家实验室 – 1/25

在核能、网络安全和科学突破方面开展合作



KUUM

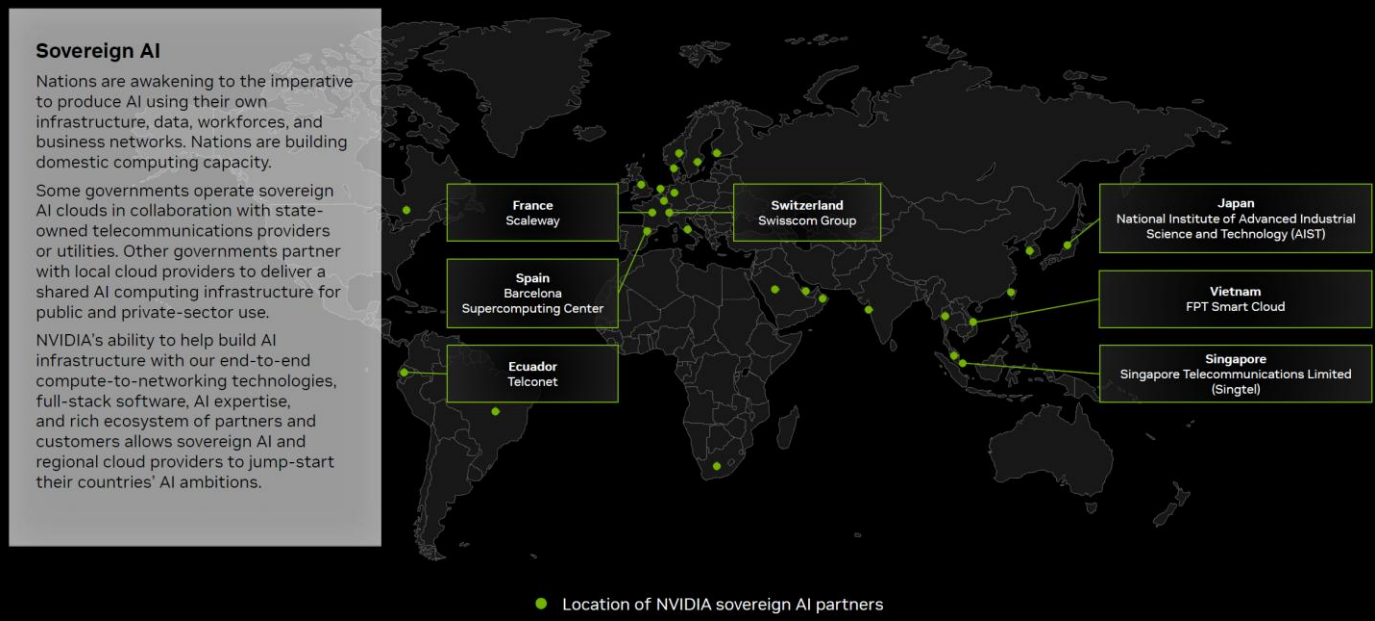
89.9 FM Live From The University of New Mexico

OpenAI touts new government partnership and support for A.I. infrastructure

# Government = Increasingly Adopting Sovereign AI Policies

政府 = 越来越多地采用主权 AI 政策

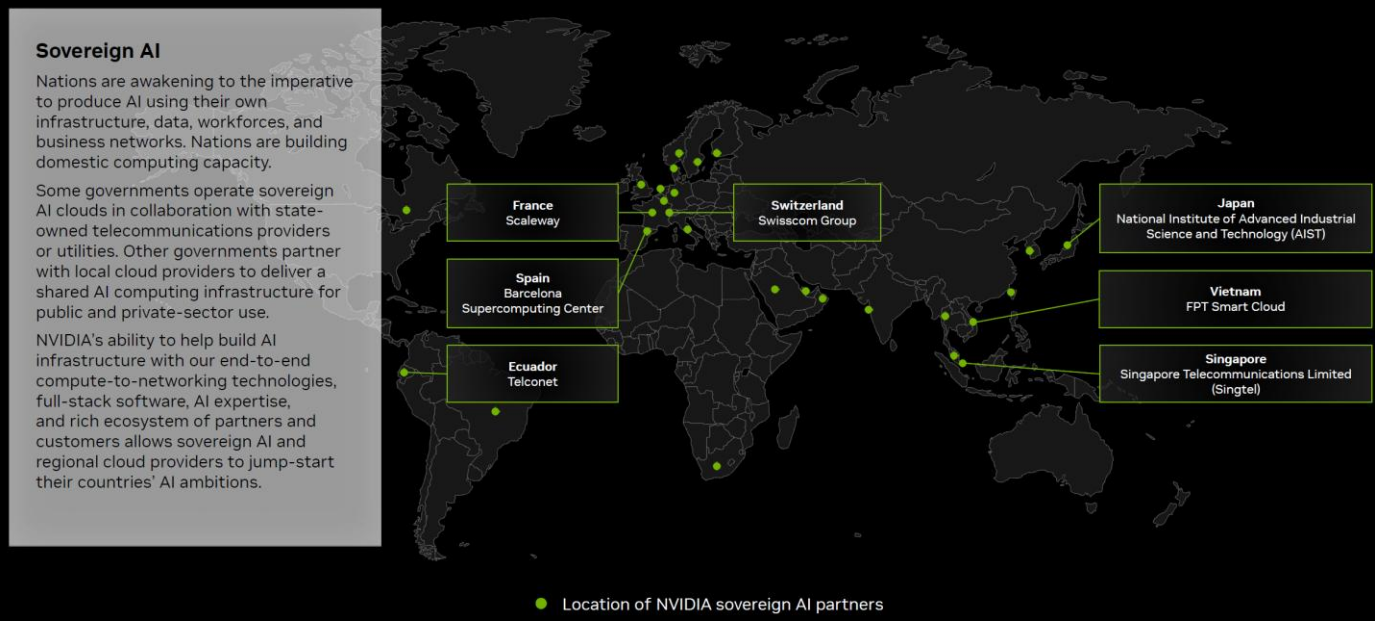
NVIDIA Sovereign AI Partners – 2/25, Per NVIDIA



*Nations are investing in AI infrastructure like they once did for electricity and Internet.*

**- NVIDIA Co-Founder & CEO Jensen Huang, 5/25**

NVIDIA Sovereign AI Partners – 2/25, Per NVIDIA

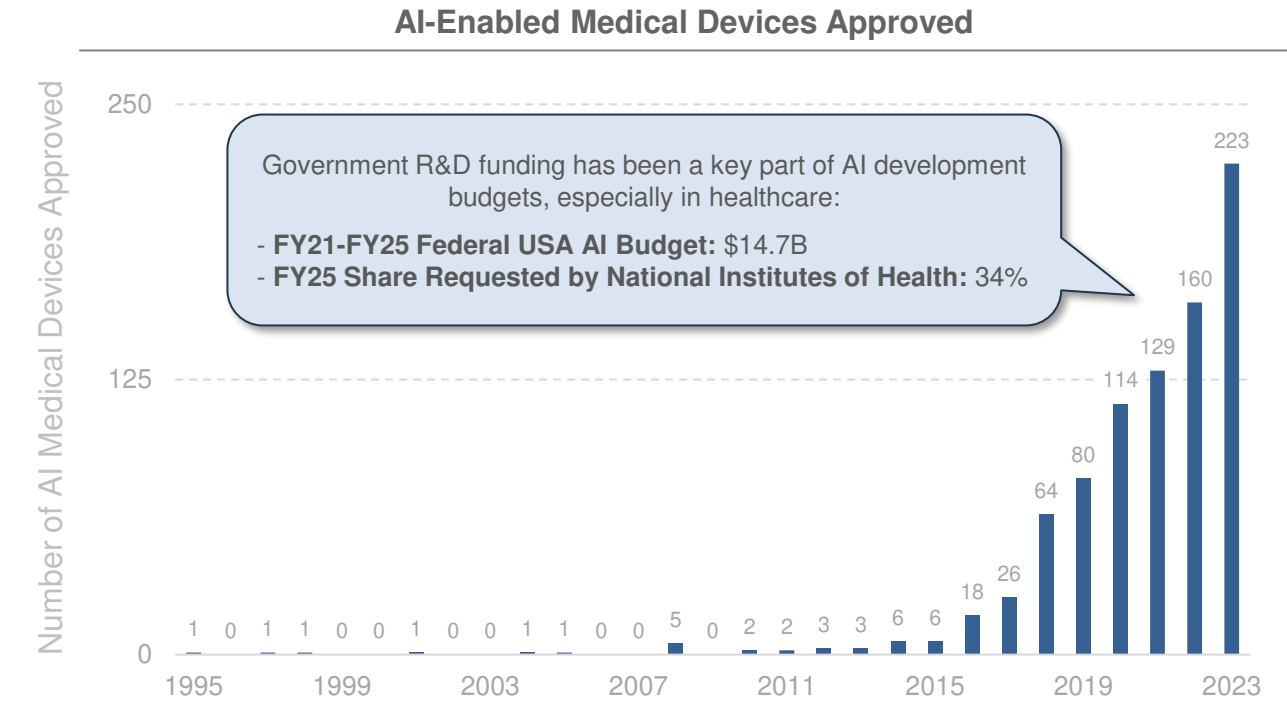


各国正在投资于 AI 基础设施, 就像他们曾经为电力和互联网所做的那样。

**- NVIDIA 联合创始人兼首席执行官 Jensen Huang, 5/25**

# Research = Rapid Ramp in FDA-Approved AI Medical Devices, per Stanford HAI

## New AI-Enabled Medical Devices Approved by USA Food & Drug Administration – 1995-2023, per Stanford HAI & USA FDA



New USA FDA AI Policy (5/25)

*In a historic first for the [USA FDA], FDA Commissioner Martin A. Makary, M.D., M.P.H., today announced an aggressive timeline to scale use of artificial intelligence (AI) internally across all FDA centers by June 30, 2025...*

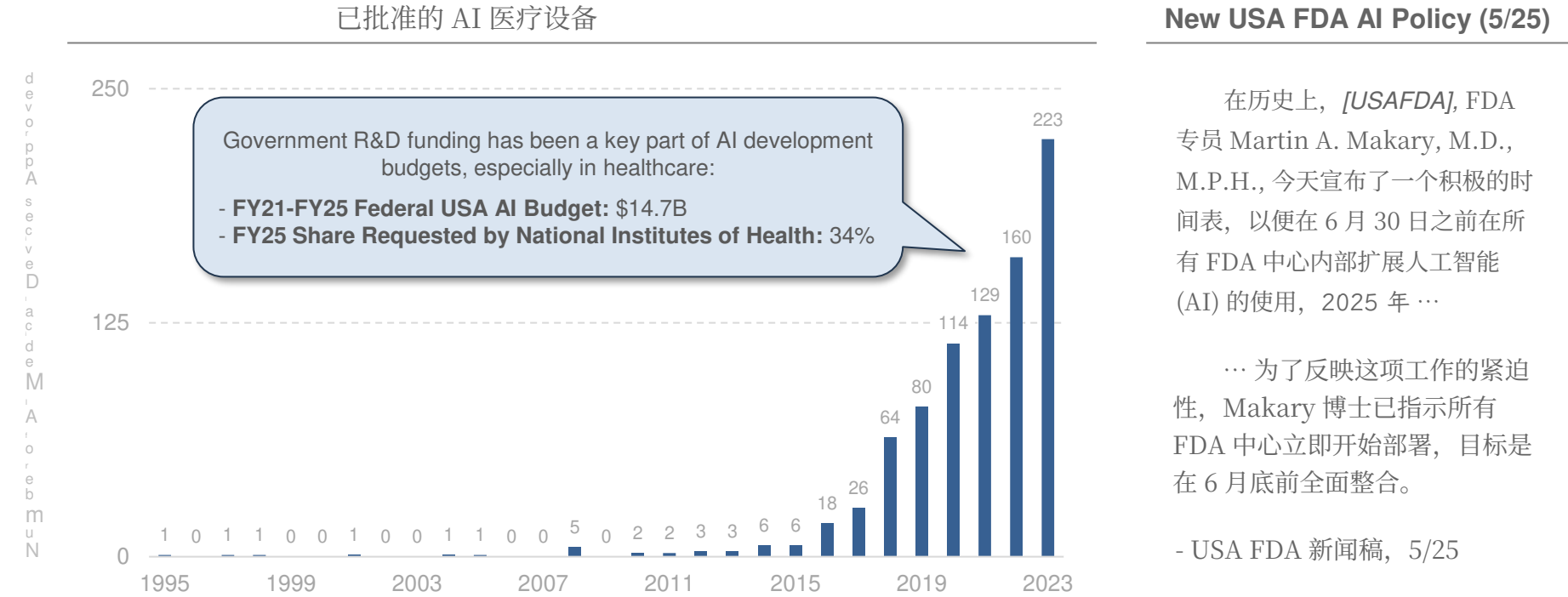
*...To reflect the urgency of this effort, Dr. Makary has directed all FDA centers to begin deployment immediately, with the goal of full integration by the end of June.*

- USA FDA Press Release, 5/25

Note: FY21, FY22 & FY23 USA government budget figures are actuals. FY24 data is enacted but not actual, FY25 data is requested. NIH share of total budget is requested.  
Source: Nestor Maslej et al., 'The AI Index 2025 Annual Report,' AI Index Steering Committee, Stanford HAI (4/25); USA Food & Drug Administration, 'FDA Announces Completion of First AI-Assisted Scientific Review Pilot and Aggressive Agency-Wide AI Rollout Timeline' (5/25); NITRD.gov (5/25)

# Research = Rapid Ramp in FDA-Approved AI Medical Devices, per Stanford HAI

## 美国食品和药物管理局批准的新型人工智能医疗设备 –1995-2023， 根据 Stanford HAI 和 USA FDA

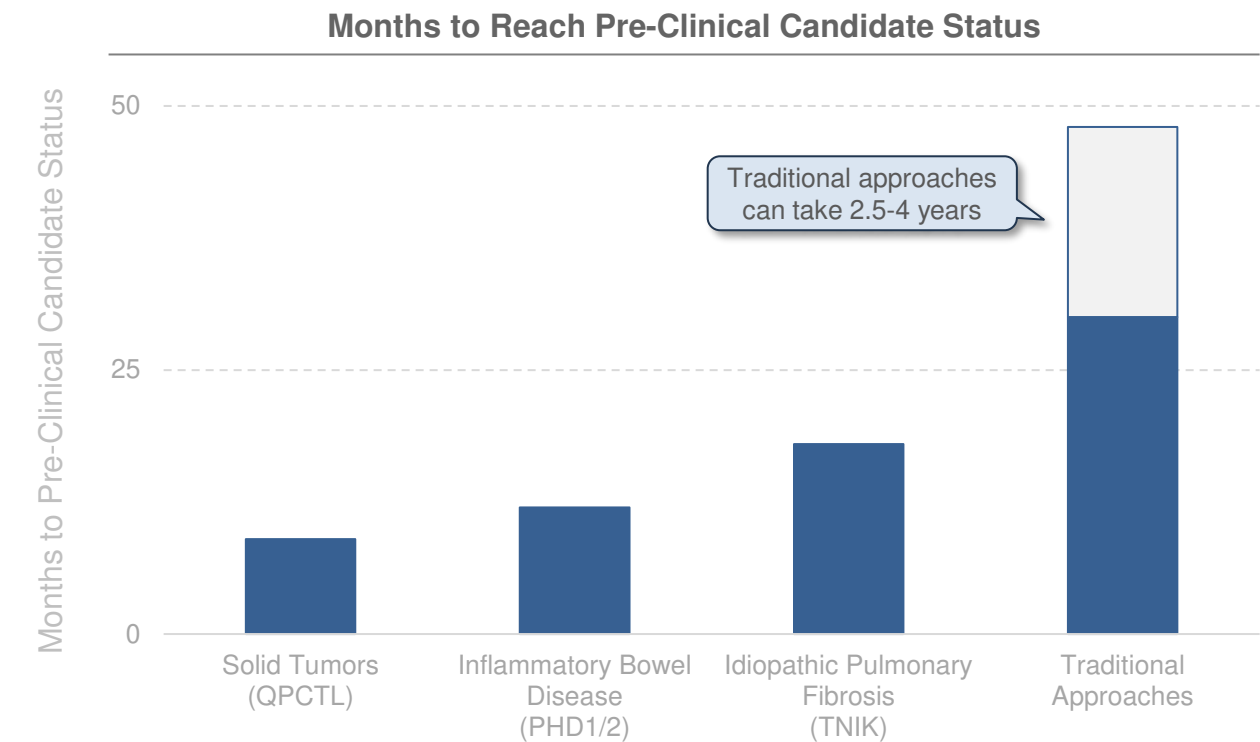



注意: FY21、FY22 和 FY23 美国政府预算数字是实际数字。FY24 数据已颁布但不是实际数据, FY25 数据是请求数据。NIH 在总预算中的份额是请求的。来源: Nestor Maslej et al., 'The AI Index 2025 Annual Report,' AI Index Steering Committee, Stanford HAI (4/25); USA Food & Drug Administration, 'FDA Announces Completion of First AI-Assisted Scientific Review Pilot and Aggressive Agency-Wide AI Rollout Timeline' (5/25); NITRD.gov (5/25)

Research =  
30%-80% Reduction in Medical R&D Timelines, per Insilico Medicine & Cradle

Research =  
30%-80% Reduction in Medical R&D Timelines, per Insilico Medicine & Cradle

AI-Driven Drug Discovery – 2021-2024, Per Insilico Medicine, Cradle & BioPharmaTrend



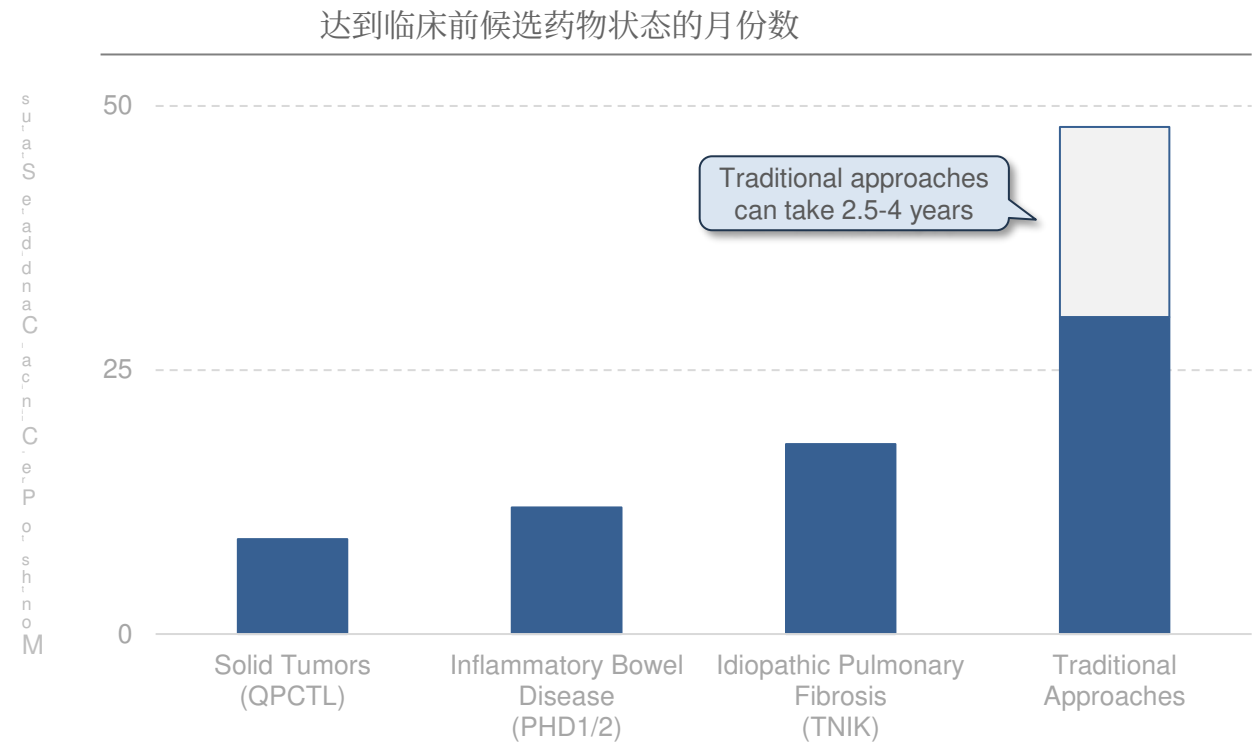
 **Cradle**


*Pharma companies that use Cradle are seeing a 1.5x to 12x speedup in pre-clinical research and development by using our GenAI platform to engineer biologics.*

**- Stef van Grieken, Co-Founder & CEO of Cradle, 5/25**

Note: Pre-Clinical Candidate Status marks the point at which a lead molecule (or biologic) has satisfied all discovery-stage gates and is officially handed off to the development organization for work related to beginning human clinical trials. Figures collected from 2021-2024. Source: Cradle, Insilico Medicine via BioPharmaTrend, 'Insilico Medicine Reports Benchmarks for its AI-Designed Therapeutics' (2/25)

AI-Driven Drug Discovery – 2021-2024, Per Insilico Medicine, Cradle & BioPharmaTrend



 **Cradle**

使用 Cradle 的制药公司通过使用我们的 GenAI 平台来设计生物制剂，其临床前研发速度提高了 1.5 倍到 12 倍。

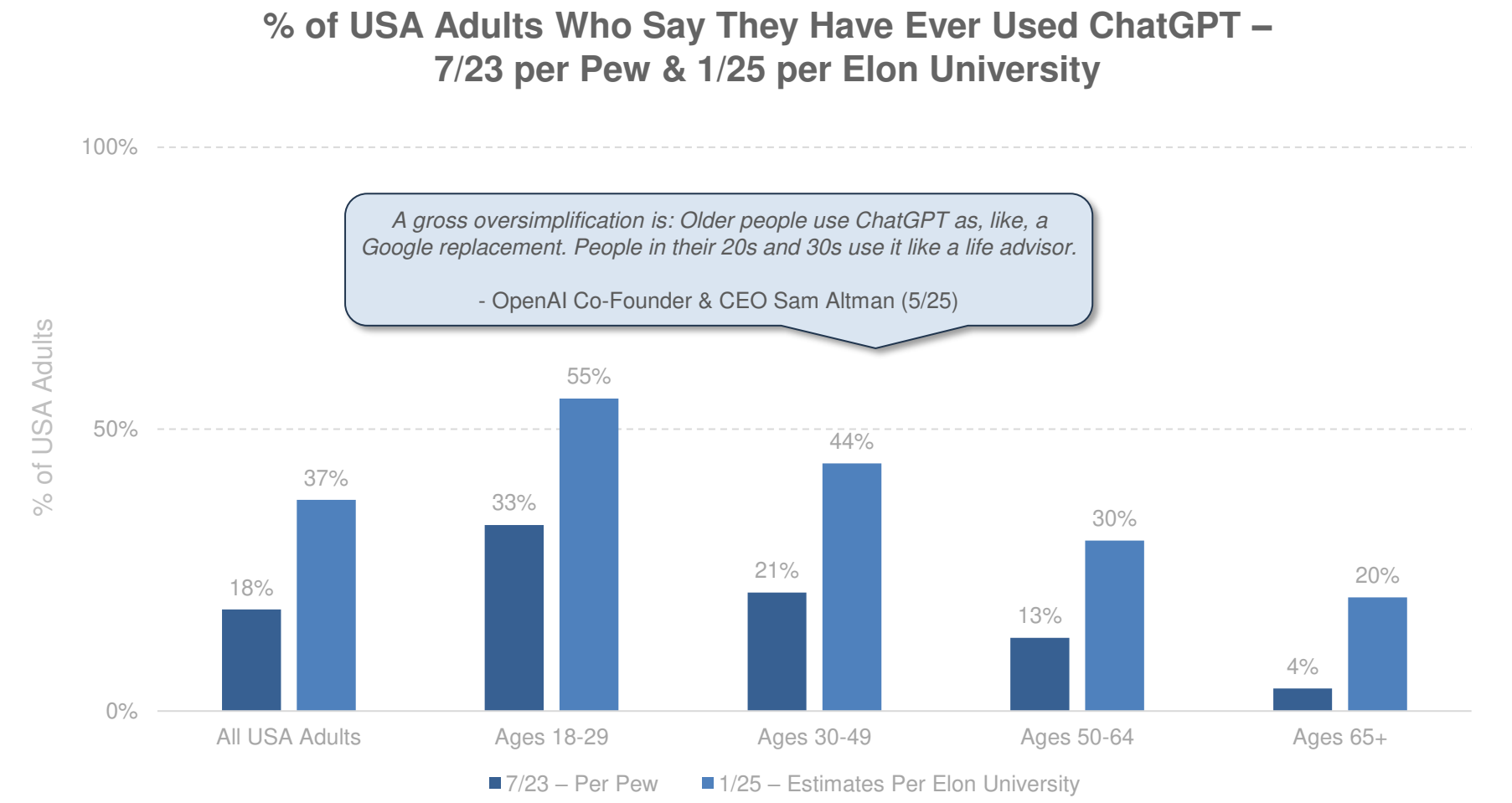
**- Stef van Grieken, Cradle 联合创始人兼首席执行官, 5/25**

注意：临床前候选药物状态标志着先导分子（或生物制剂）已满足所有发现阶段的条件，并正式移交给开发组织，以进行与开始人体临床试验相关的工作。数据收集自 2021-2024 年。来源：Cradle, Insilico Medicine via BioPharmaTrend, 'Insilico Medicine Reports Benchmarks for its AI-Designed Therapeutics' (2/25)

*AI User + Usage + CapEx Growth =*  
*Unprecedented*

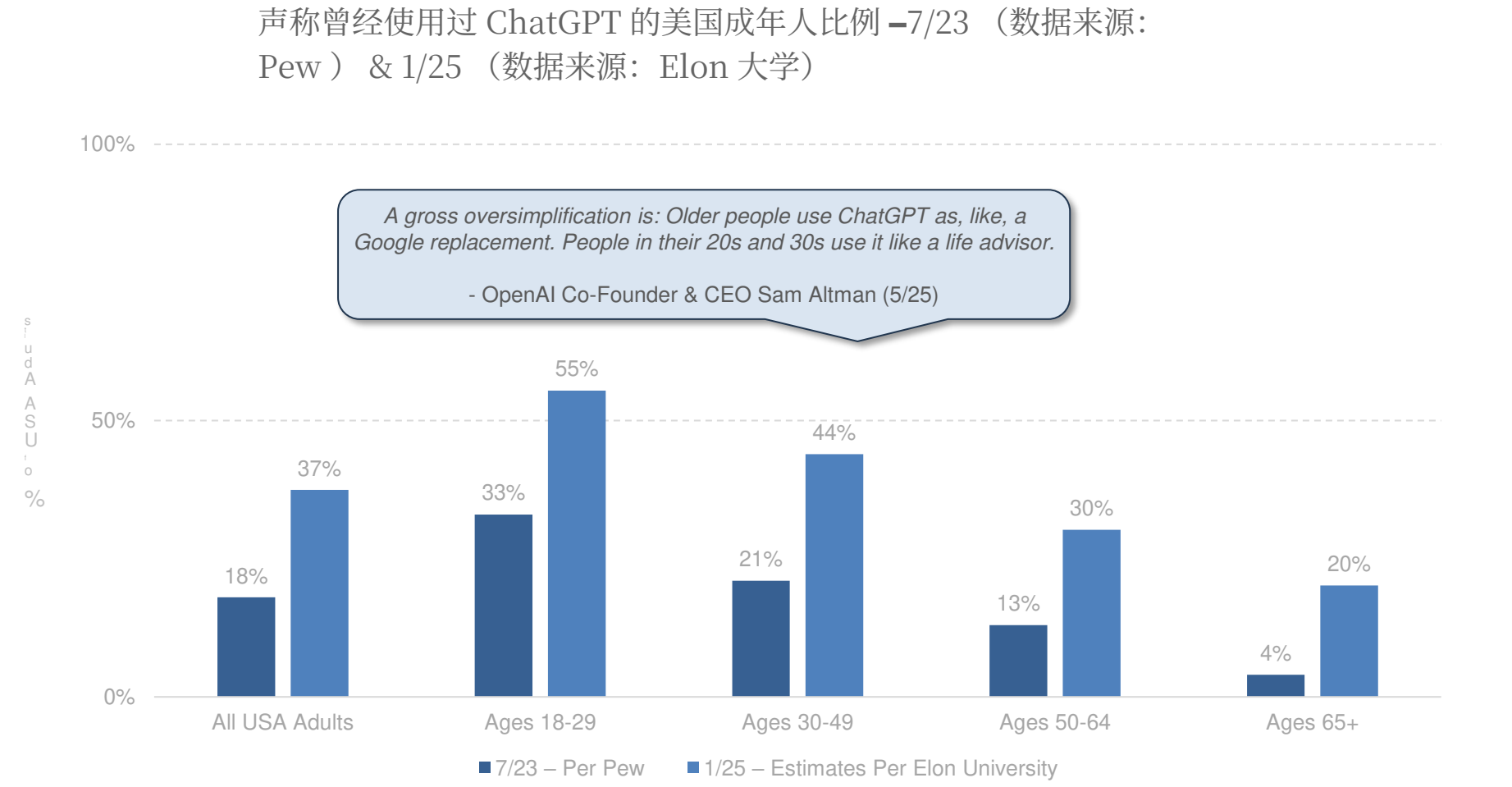
AI 用户 + 使用量 + 资本支出增长 =  
前所未有

# AI Usage – ChatGPT = Rising Rapidly Across Age Groups in USA, per Pew & Elon University



*Note: 7/23 data per Pew Research study on ChatGPT use, n=10,133 USA adults. Those who did not give an answer are not shown. 1/25 data per Elon University study on use of any AI models, n=500 USA adults,. Figures estimated based on overall AI tool usage adjusted for an average 72% usage rate of ChatGPT amongst respondents who use any AI tools. Actual ChatGPT penetration may vary by cohort. Note that this chart aggregates data across survey providers and as such may not be directly comparable. Source: Pew Research Center (3/26/24), Elon University (released 3/12/25), Sam Altman (5/12/25) via Fortune*

# AI 使用情况 – ChatGPT = 在美国各年龄段人群中迅速增长，数据来源：Pew 和 Elon 大学

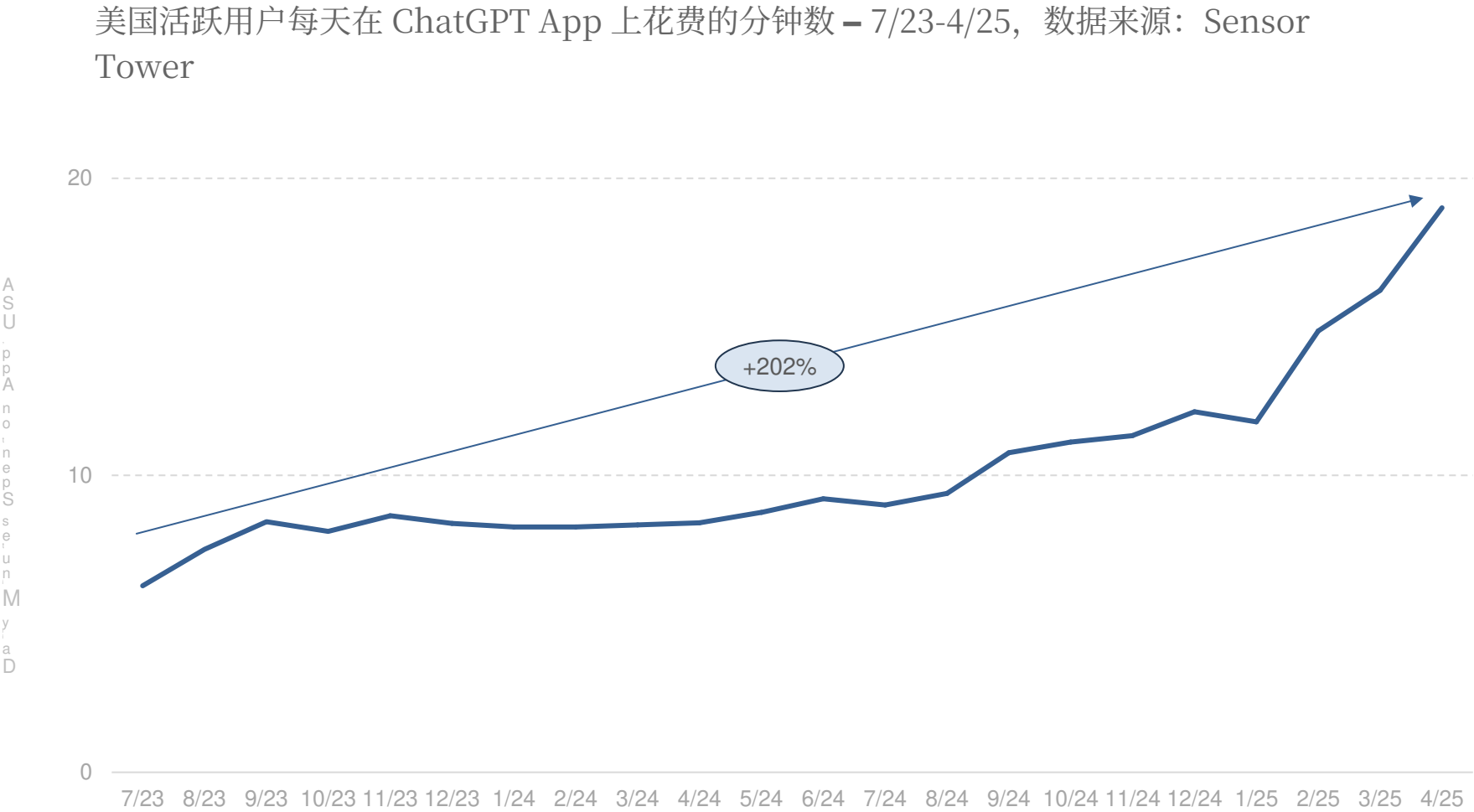
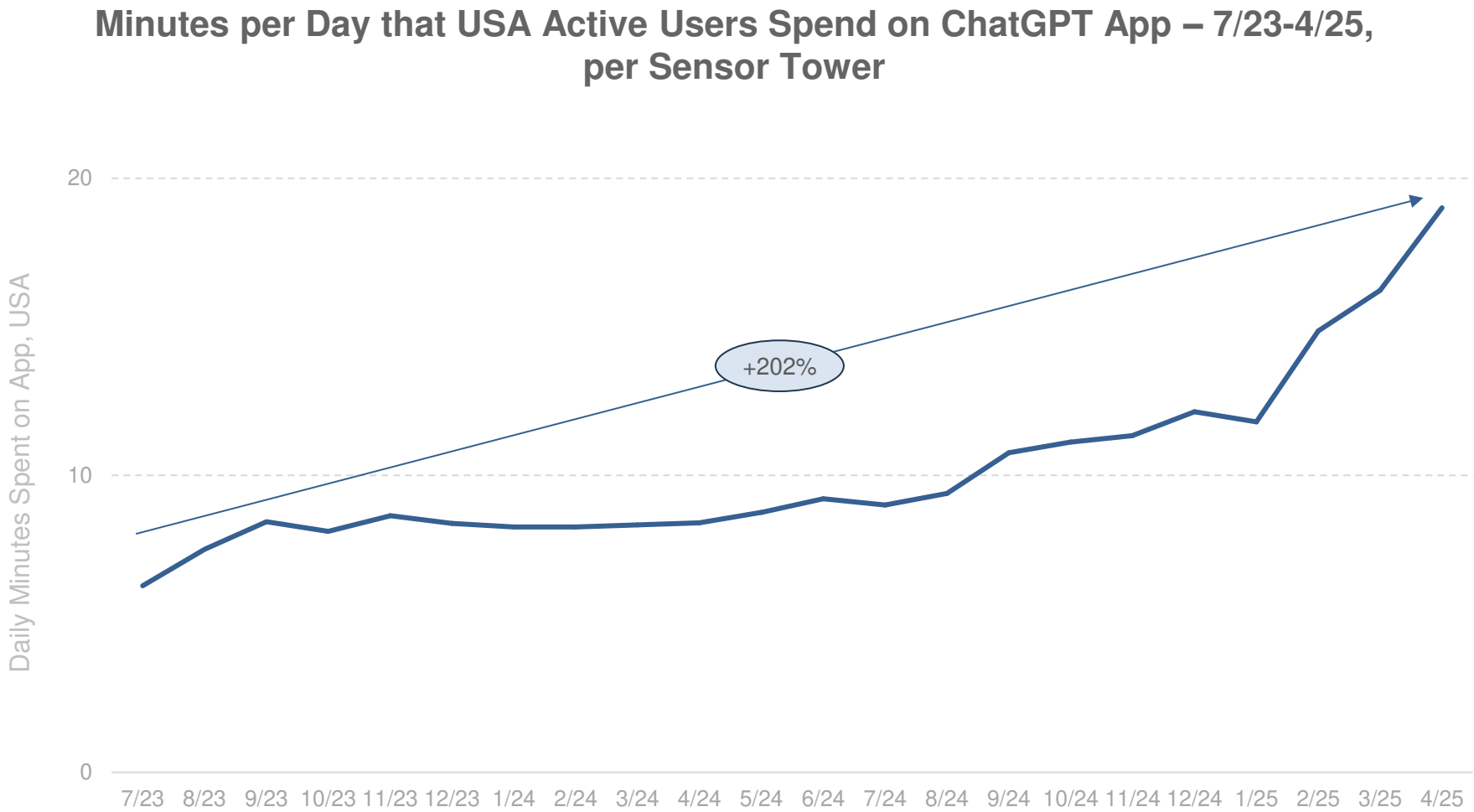


注：7/23 数据来自 Pew Research 关于 ChatGPT 使用情况的研究，n=10,133 名美国成年人。未给出答案的人未显示。1/25 数据来自 Elon 大学关于任何 AI 模型的使用情况的研究，n=500 名美国成年人。数据基于对任何 AI 工具的使用情况进行调整后的总体 AI 工具使用率（ ChatGPT 在受访者中平均使用率为 72% ）进行估算。实际的 ChatGPT 普及率可能因人而异。请注意，此图表汇总了来自不同调查提供商的数据，因此可能不具有直接可比性。来源：Pew Research Center (2024 年 3 月 26 日)、Elon University (2025 年 3 月 12 日发布)、Sam Altman (2025 年 5 月 12 日) 通过 Fortune



AI Engagement (ChatGPT App as Proxy) = +202% Rise in Daily Time Spent Over Twenty-One Months...

AI 参与度（以 ChatGPT App 为代表） =+202% 每日花费时间在二十个月内增长 ...

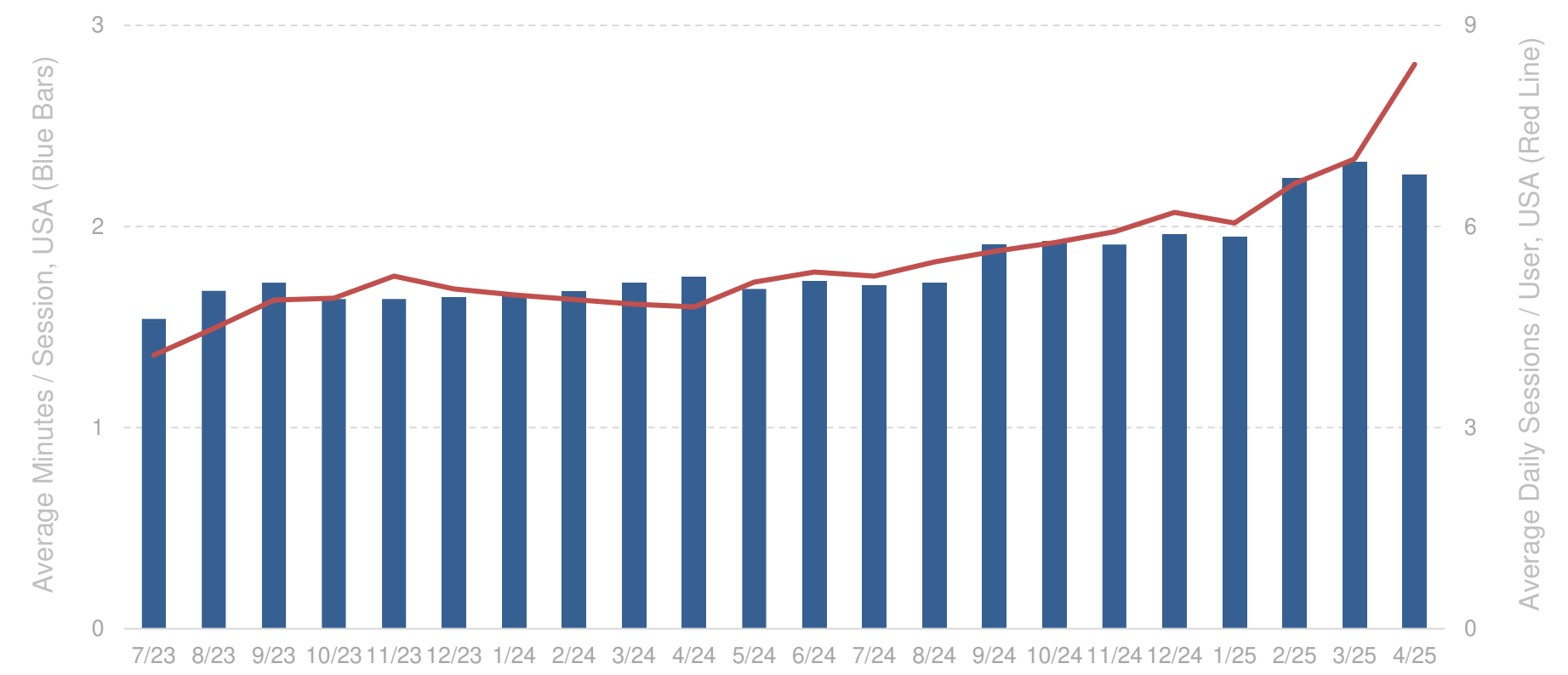


Note: Data represents USA App Store & Google Play Store monthly active users. Data for ChatGPT standalone app only. ChatGPT app not available in China, Russia and select other countries as of 5/25. Source: Sensor Tower (5/25)

注意：数据代表美国 App Store 和 Google Play 商店的月活跃用户。数据仅适用于 ChatGPT 独立应用程序。截至 5 月 25 日，ChatGPT 应用程序在中国、俄罗斯和部分其他国家 / 地区不可用。来源：Sensor Tower (5/25)

...AI Engagement (ChatGPT App as Proxy) = +106% Growth in Sessions & +47% Growth in Duration Over Twenty-One Months

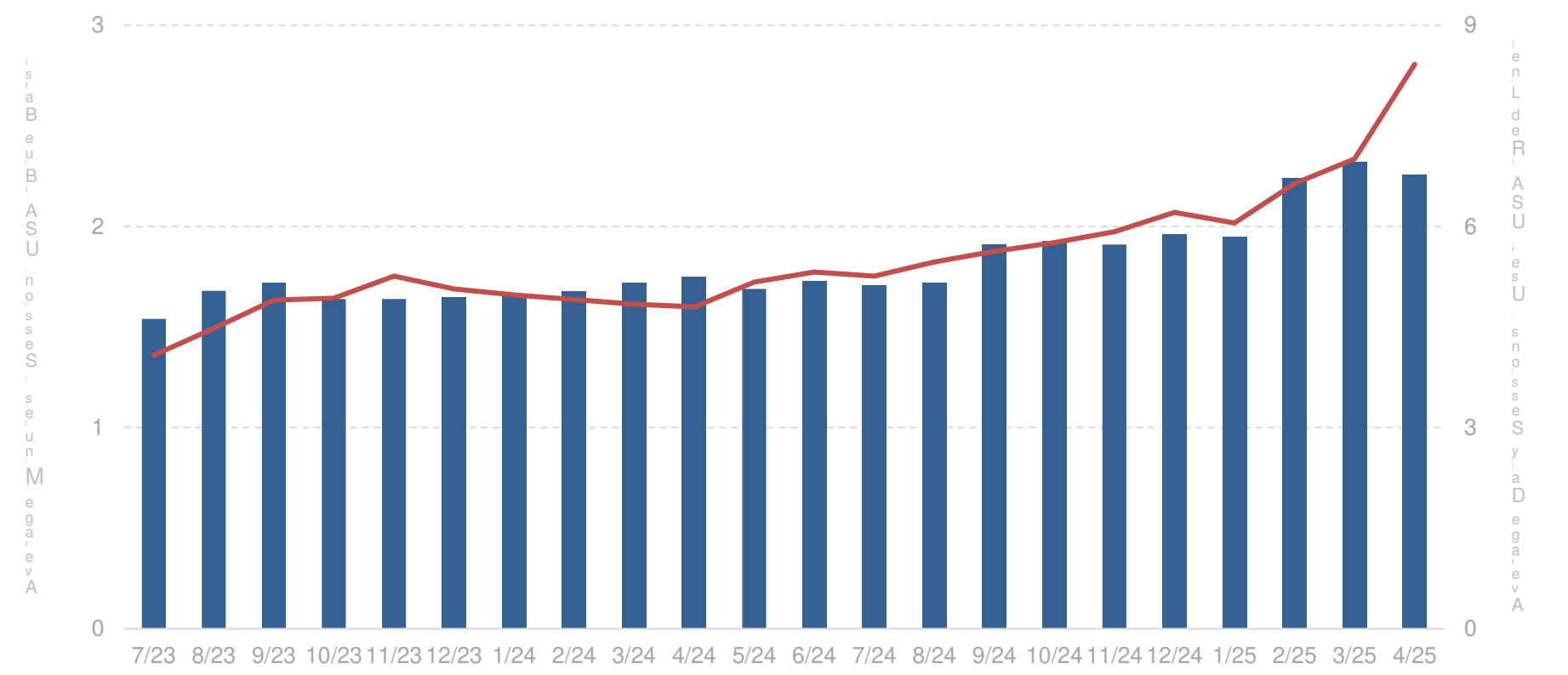
Average USA Session Duration (Minutes) & Daily Sessions per User for ChatGPT App – 7/23-4/25, per Sensor Tower



Note: Data represents USA App Store & Google Play Store monthly active users. Data for ChatGPT standalone app only. ChatGPT app not available in China, Russia and select other countries as of 5/25. Source: Sensor Tower (5/25)

...AI Engagement (ChatGPT App as Proxy) =+106% Sessions 增长 & +47% 持续时间在二十一个月内增长

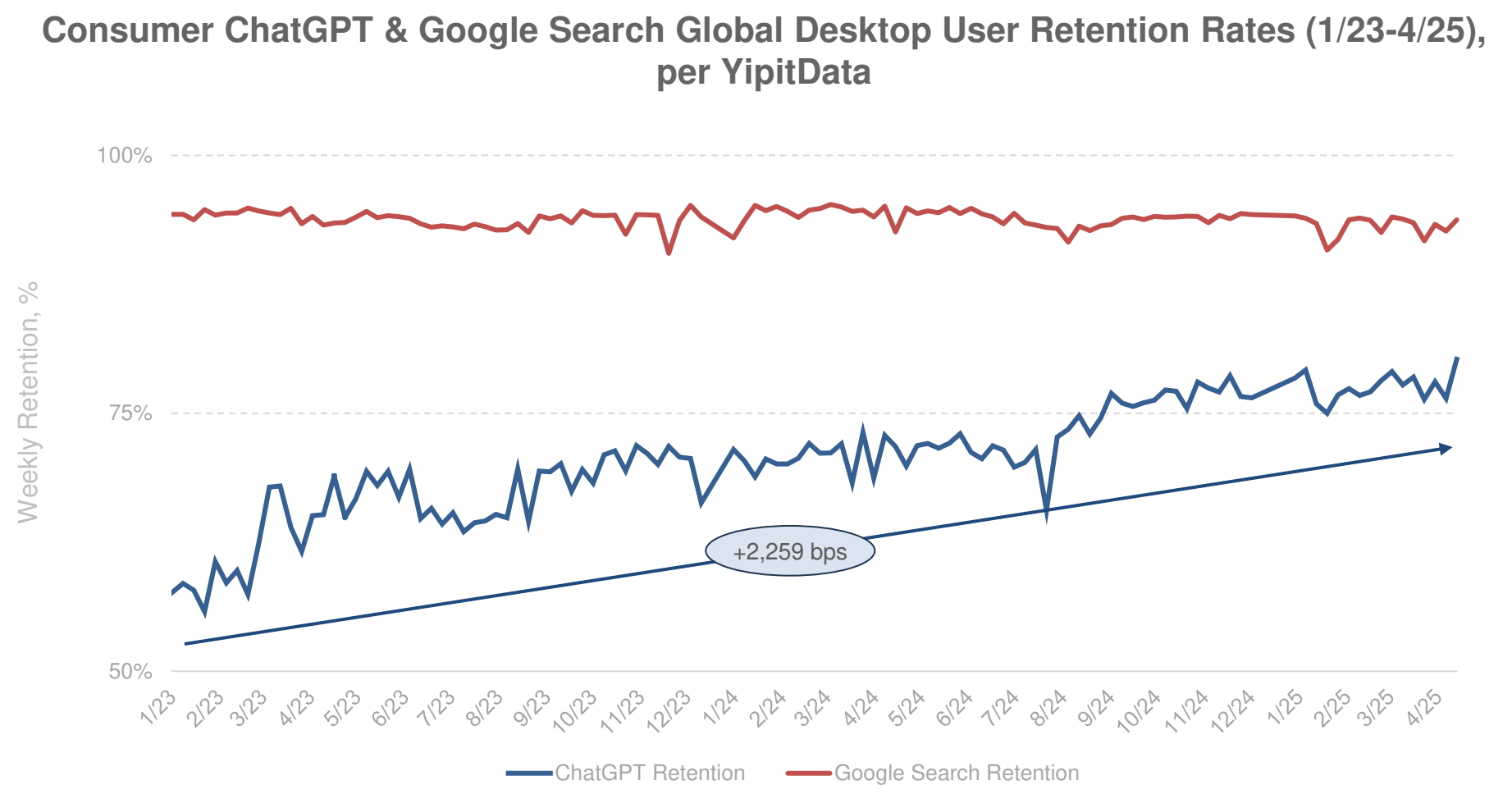
ChatGPT 应用程序的美国平均会话时长（分钟）和每个用户的每日会话次数 –7/23-4/25，数据来源：Sensor Tower



注意：数据代表美国 App Store 和 Google Play 商店的月活跃用户。数据仅适用于 ChatGPT 独立应用程序。截至 5 月 25 日，ChatGPT 应用程序在中国、俄罗斯和部分其他国家 / 地区不可用。来源：Sensor Tower (5/25)

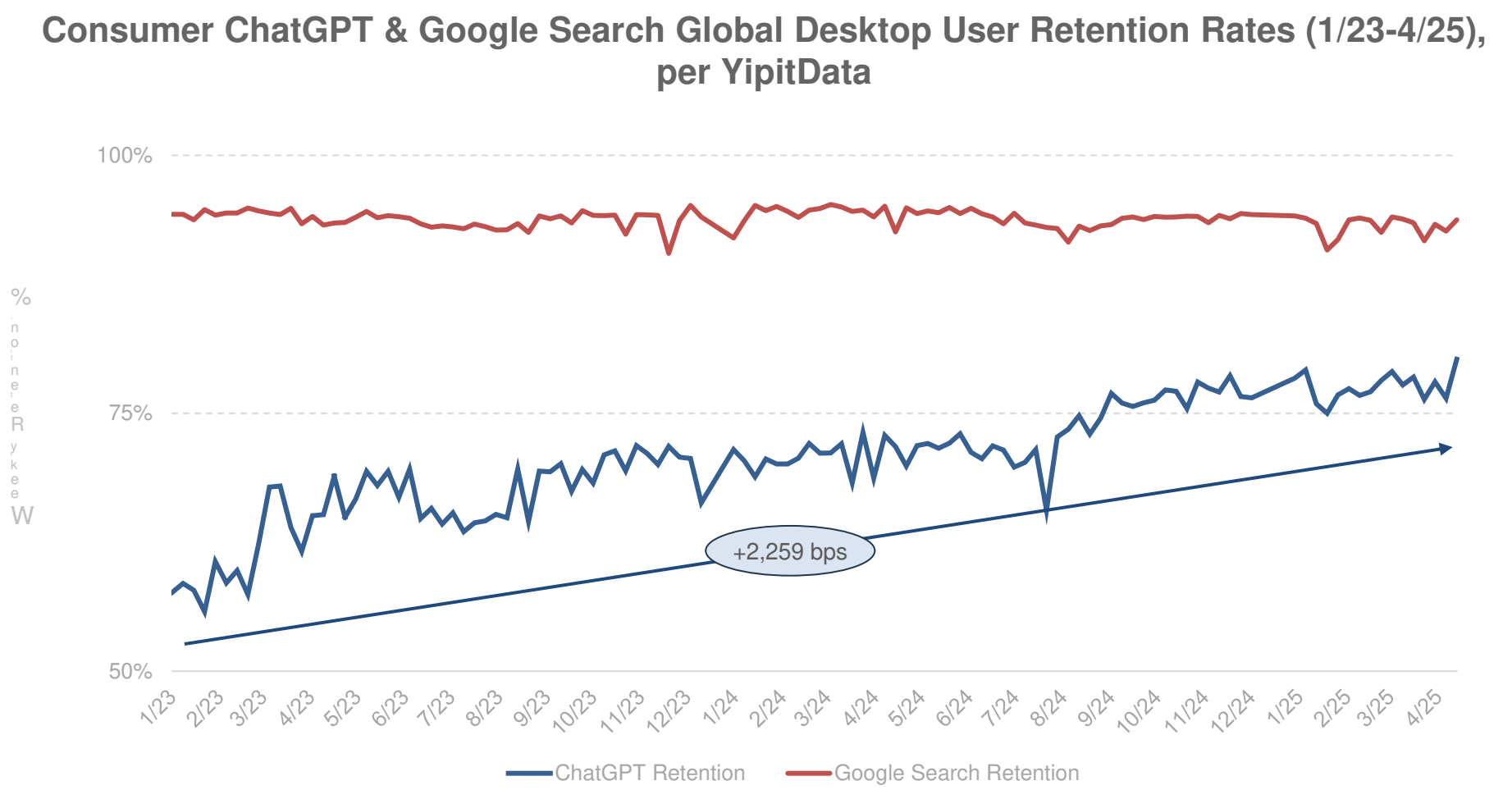


# AI Retention (ChatGPT as Proxy) = 80% vs. 58% Over Twenty-Seven Months, per YipitData



Note: Retention Rate = Percentage of users from the immediately preceding week that were users again in the current week. Data measures several million global active desktop users' clickstream data. Data consists of users' web requests & is collected from web services / applications, such as VPNs and browser extensions. Users must have been part of the panel for 2 consecutive months to be included. Panel is globally-representative, though China data may be subject to informational limitations due to government restrictions. Excludes anomalies in w/c 12/24/23, 12/31/23, 12/22/24, 12/29/24, 1/5/25, potentially due to holiday breaks causing less enterprise usage. Source: YipitData (5/25)

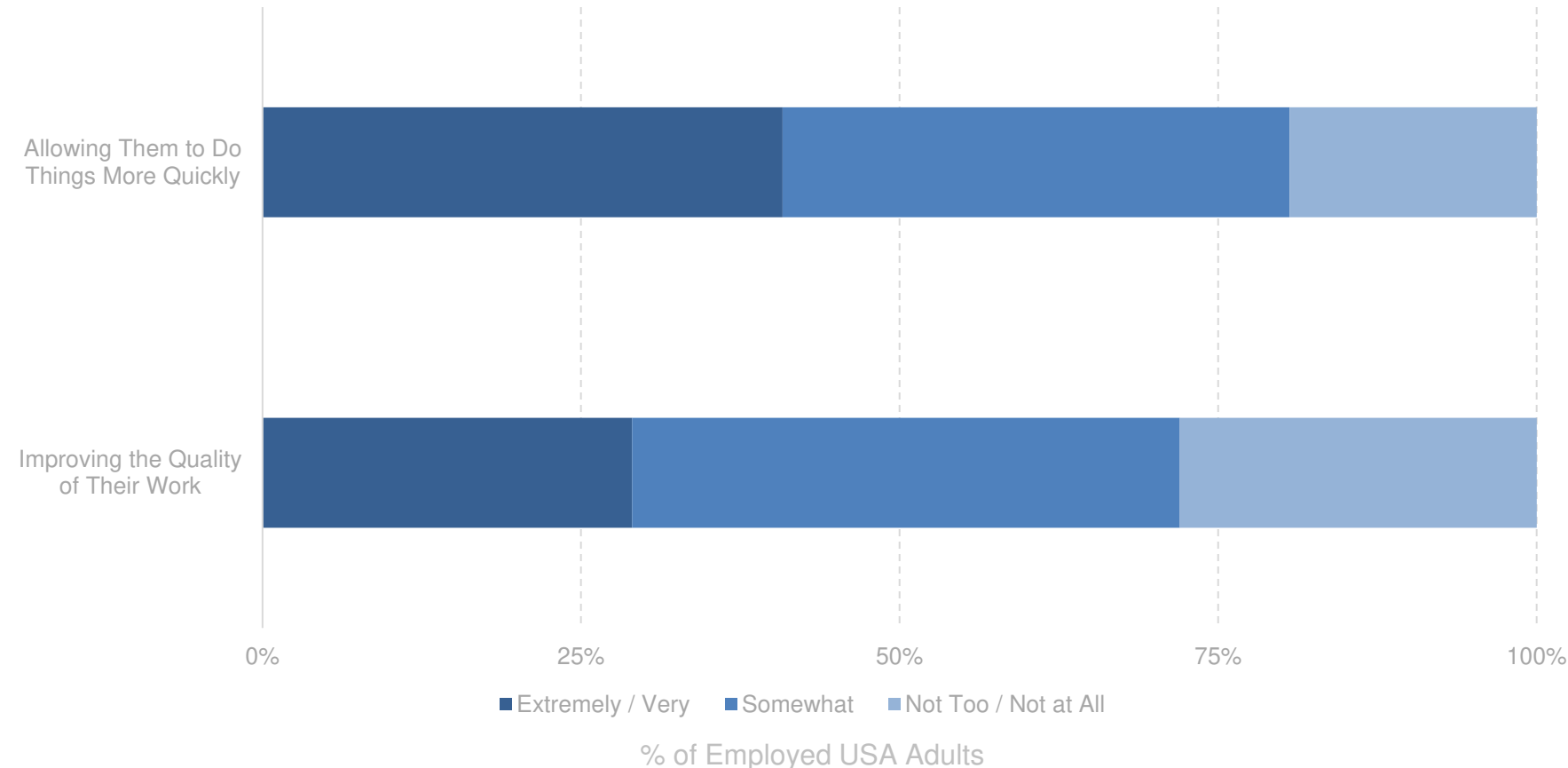
# AI 用户留存率（以 ChatGPT 作为参考）=80% vs. 58%，时间跨度为 27 个月（数据来源：YipitData）



注意：留存率 = 指前一周的用户在本周再次成为用户的百分比。数据衡量了数百万全球活跃桌面用户的点击流数据。数据由用户的网络请求组成，并从网络服务 / 应用程序（如 VPN 和浏览器扩展）收集。用户必须连续 2 个月成为面板的一部分才能被纳入。面板具有全球代表性，但由于政府限制，中国的数据可能受到信息限制。排除 12/24/23、12/31/23、12/22/24、12/29/24、1/5/25 当周的异常情况，这可能是由于假期导致企业使用量减少。来源：YipitData（5/25）

AI Chatbots @ Work Tells =  
>72% Doing Things Quicker / Better

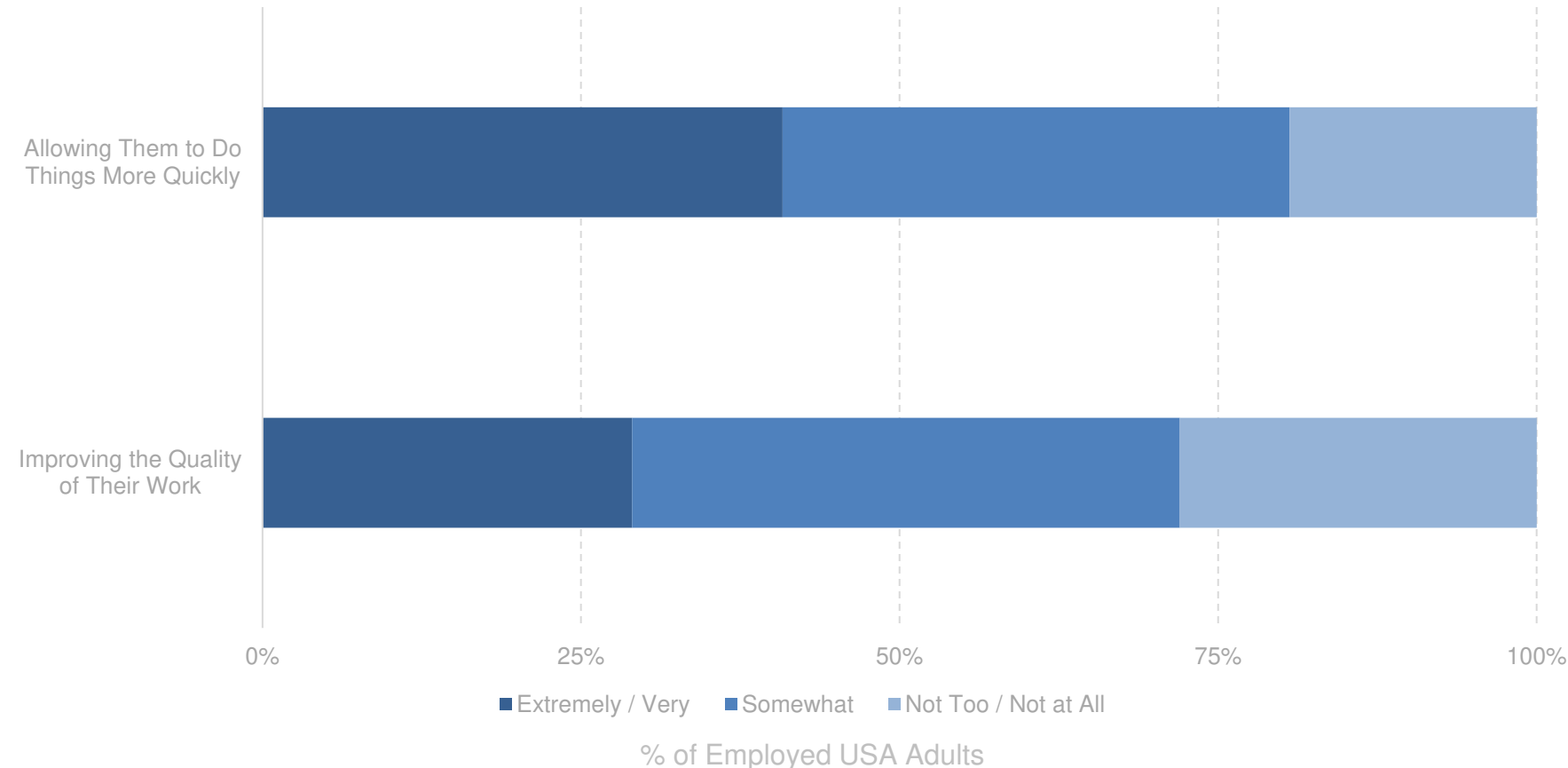
% of Employed USA Adults Using AI Chatbots Who Say Tools Have Been \_\_\_\_\_  
Helpful When It Comes to... – 10/24, per Pew



Note: N = 5,273 USA adults who are employed part time or full time and who have only one job or have more than one but consider one of them to be their primary job were surveyed.  
Source: Pew Research Center (10/24)

AI Chatbots @ Work Tells =>72% 做  
事更快 / 更好

使用 AI Chatbots 的美国成年人中，表示这些工具在 \_\_\_\_\_ 方面有帮助的人数百分比在以下方  
面有所帮助 … – 10/24，每皮尤

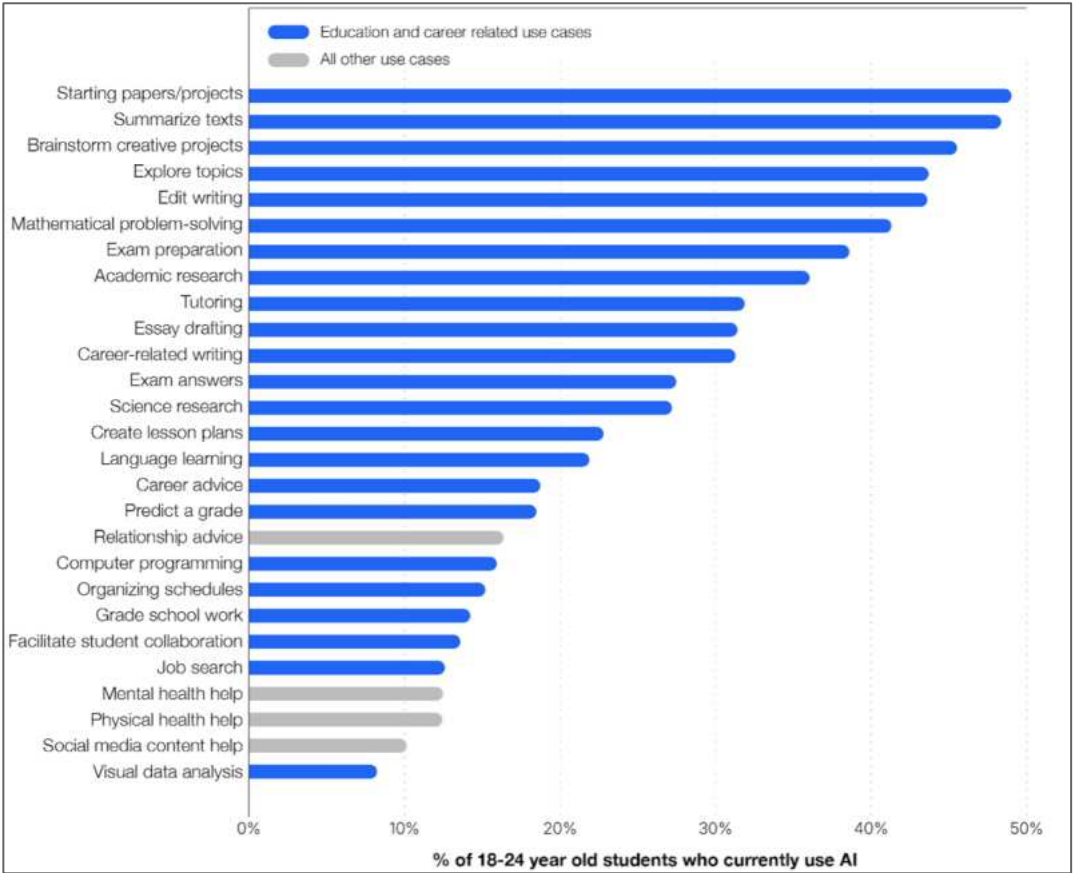


注意：调查对象为 N = 5,273 名从事兼职或全职工作且只有一份工作或多份工作但认为其中一份是主要工作的美国成年人。来源：皮尤研究中心 (10/24)

# AI Chatbots @ School Tells (ChatGPT as Proxy) = Bias to Research / Problem Solving / Learning / Advice

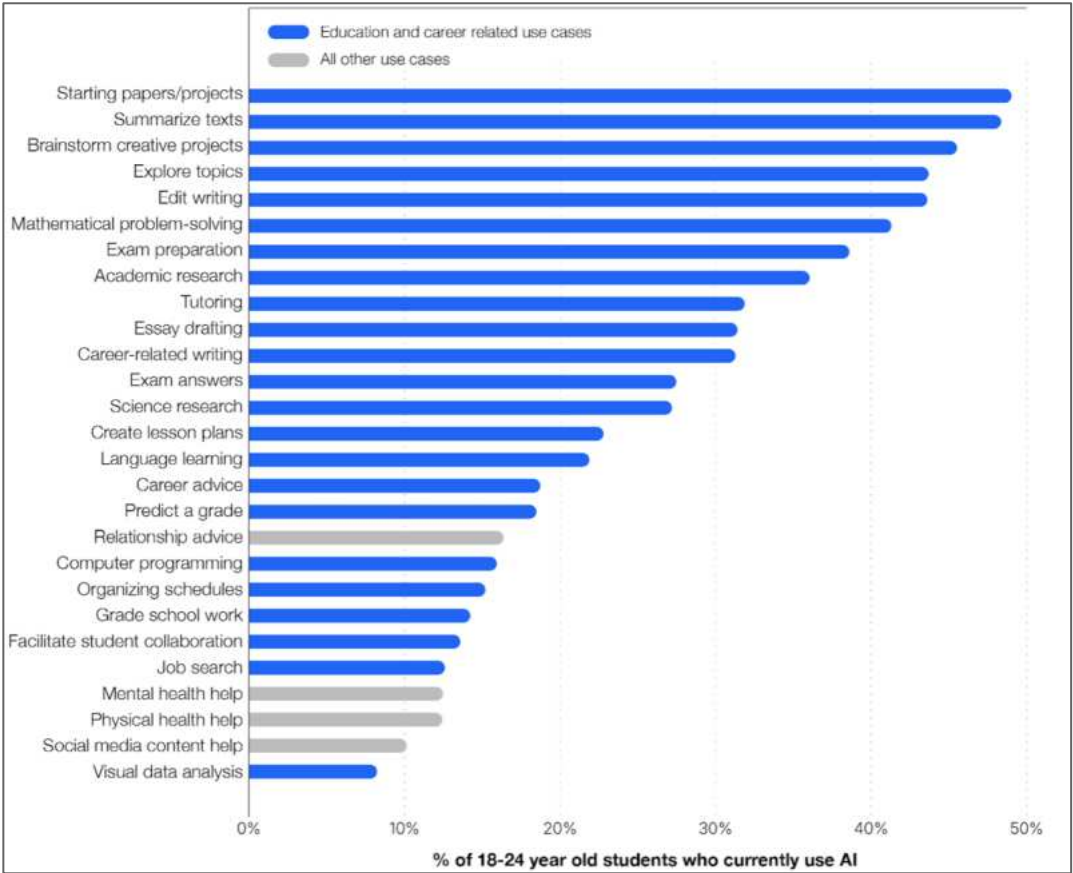
# 学校中的人工智能聊天机器人讲述（ ChatGPT 作为代理） = 对研究 / 问题解决 / 学习 / 建议的偏见

OpenAI ChatGPT Usage Survey, USA Students Ages 18-24 – 12/24-1/25, per OpenAI



Note: Data per OpenAI survey (12/24), n = 1,299 USA college and graduate students across a mix of STEM and non-STEM disciplines; only answers from 18-24 year olds used. Sample includes both AI users and non-users but excludes “AI rejectors” – defined as non-users with little to no interest in adopting AI within the next 12 months. Source: OpenAI, ‘Building an AI-Ready Workforce: A Look at College Student ChatGPT Adoption in the US’ (2/25)

OpenAI ChatGPT Usage Survey, USA Students Ages 18-24 – 12/24-1/25, per OpenAI



注意：数据来自 OpenAI 调查 (12/24)，n = 1,299 名美国大学生和研究生，涵盖 STEM 和非 STEM 学科的组合；仅使用 18-24 岁年龄段的答案。样本包括人工智能用户和非用户，但不包括 “人工智能拒绝者”—— 定义为未来 12 个月内对采用人工智能几乎没有兴趣的非用户。来源：OpenAI, ‘构建人工智能 - 就绪的劳动力：美国大学生 ChatGPT 采用情况分析’ (2/25)

AI Usage Expansion – Deep Research = Automating Specialized Knowledge Work

AI 使用扩展 – 深度研究 = 自动化专业知识工作

Select AI Company Deep Research Capabilities – 12/24-2/25, per Google, OpenAI & xAI

Select AI Company Deep Research Capabilities – 12/24-2/25, per Google, OpenAI & xAI

Google Gemini Deep Research

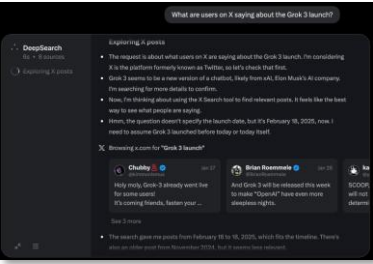
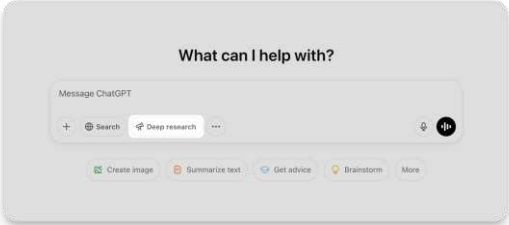
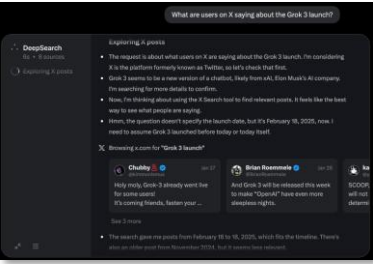
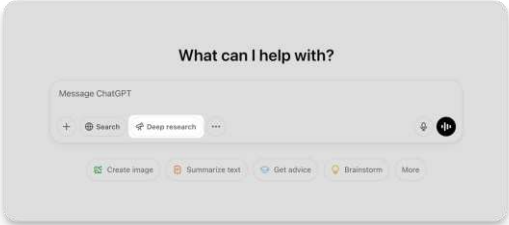
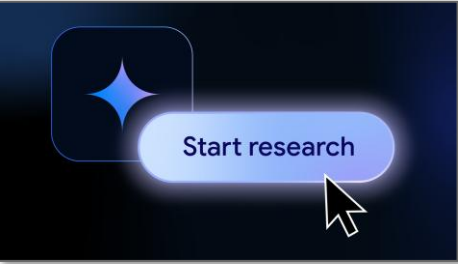
OpenAI ChatGPT Deep Research

xAI Grok DeepSearch

Google Gemini 深度研究

OpenAI ChatGPT 深度研究

xAI Grok DeepSearch



Get up to speed on just about anything with Deep Research, an agentic feature in Gemini that can automatically browse up to hundreds of websites on your behalf, think through its findings, and **create insightful multi-page, reports that you can turn into engaging podcast-style conversations...**

...It's a step towards more agentic AI that can move beyond simple question-answering to become a true collaborative partner.

- Google Deep Research Overview, launched 12/24

Today we're launching deep research in ChatGPT, a new agentic capability that conducts multi-step research on the internet for complex tasks. **It accomplishes in tens of minutes what would take a human many hours...**

...Deep research marks a significant step toward our broader goal of developing AGI, which we have long envisioned as capable of producing novel scientific research.

- OpenAI Deep Research Press Release, 2/25

To understand the universe, we must interface Grok with the world...

...As a first step towards this vision, we are rolling out DeepSearch – our first agent. **It's a lightning-fast AI agent built to relentlessly seek the truth across the entire corpus of human knowledge. DeepSearch is designed to synthesize key information, reason about conflicting facts and opinions, and distill clarity from complexity.**

- xAI Grok 3 Beta Press Release, 2/25

通过深度研究快速了解几乎所有内容，这是 Gemini 中的一项代理功能，可以代表您自动浏览多达数百个网站，考虑其发现，并创建富有洞察力的多页报告，您可以将其转化为引人入胜的播客式对话 ...

... 这是迈向更具代理性的 AI 的一步，它可以超越简单的问答，成为真正的协作伙伴。

- Google 深度研究概览，于 24 年 12 月推出

Today we're launching deep research in ChatGPT, a new agentic capability that conducts multi-step research on the internet for complex tasks. 它在几十分钟内完成人类需要花费数小时才能完成的工作 ...

... 深度研究标志着朝着我们开发 AGI 的更广泛目标迈出了重要一步，我们一直认为 AGI 能够产生新颖的科学研究。

- OpenAI 深度研究新闻稿，2/25

为了理解宇宙，我们必须使 Grok 与世界交互 .....

... 作为实现这一愿景的第一步，我们正在推出 DeepSearch– 我们的第一个代理。它是一个闪电般快速的 AI 代理，旨在不懈地在整个人类知识体系中寻找真相。DeepSearch 旨在综合关键信息，推理相互冲突的事实和观点，并从复杂性中提炼出清晰的思路。

- xAI Grok 3 Beta 新闻稿，2/25

Source: Google (5/25), OpenAI (2/25), xAI (2/25), Digital Trends (1/25)

来源: Google (5/25), OpenAI (2/25), xAI (2/25), Digital Trends (1/25)

AI Agent Evolution =  
Chat Responses → Doing Work

AI Agent Evolution =  
Chat Responses → Doing Work

A new class of AI is now emerging – less assistant, more service provider.

What began as basic conversational interfaces may now be evolving into something far more capable.

Traditional chatbots were designed to respond to user prompts, often within rigid scripts or narrow flows. They could fetch answers, summarize text, or mimic conversation – but always in a reactive, limited frame.

**AI agents** represent a step-change forward. These are intelligent long-running processes that can reason, act, and complete multi-step tasks on a user’s behalf. They don’t just answer questions – they execute: booking meetings, submitting reports, logging into tools, or orchestrating workflows across platforms, often using natural language as their command layer.

This shift mirrors a broader historical pattern in technology.

Just as the early 2000s saw static websites give way to dynamic web applications – where tools like Gmail and Google Maps transformed the internet from a collection of pages into a set of utilities – AI agents are turning conversational interfaces into functional infrastructure.

Whereas early assistants needed clear inputs and produced narrow outputs, agents promise to operate with goals, autonomy and certain guardrails. They promise to interpret intent, manage memory, and coordinate across apps to get real work done. It’s less about responding and more about accomplishing.

While we are early in the development of these agents, the implications are just starting to emerge.

AI agents could reshape how users interact with digital systems – from customer support and onboarding to research, scheduling, and internal operations.

Enterprises are leading the charge; they’re not just experimenting with agents, but deploying them, investing in frameworks and building ecosystems around autonomous execution.

What was once a messaging interface is becoming an action layer.

A new class of AI is now emerging – less assistant, more service provider.

What began as basic conversational interfaces may now be evolving into something far more capable.

传统聊天机器人旨在响应用户提示，通常在严格的脚本或狭窄的流程中。它们可以获取答案、总结文本或模仿对话 – 但始终在被动的、有限的框架内。

AI 代理代表着向前迈出的一大步。这些是智能的、长期运行的流程，可以推理、行动并代表用户完成多步骤任务。它们不只是回答问题 – 它们会执行：预订会议、提交报告、登录工具或协调跨平台的 Workflow，通常使用自然语言作为其命令层。

这种转变反映了技术领域更广泛的历史模式。正如 2000 年代初期静态网站让位于动态 Web 应用程序一样 –

Gmail 和 Google Maps 等工具将互联网从页面集合转变为实用程序集合 – AI 代理正在将对话界面转变为功能性基础设施。

早期的助手需要清晰的输入并产生狭窄的输出，而代理则有望在目标、自主性和某些防护措施下运行。它们承诺解释意图、管理内存并在应用程序之间进行协调以完成实际工作。这更多的是关于完成而不是响应。

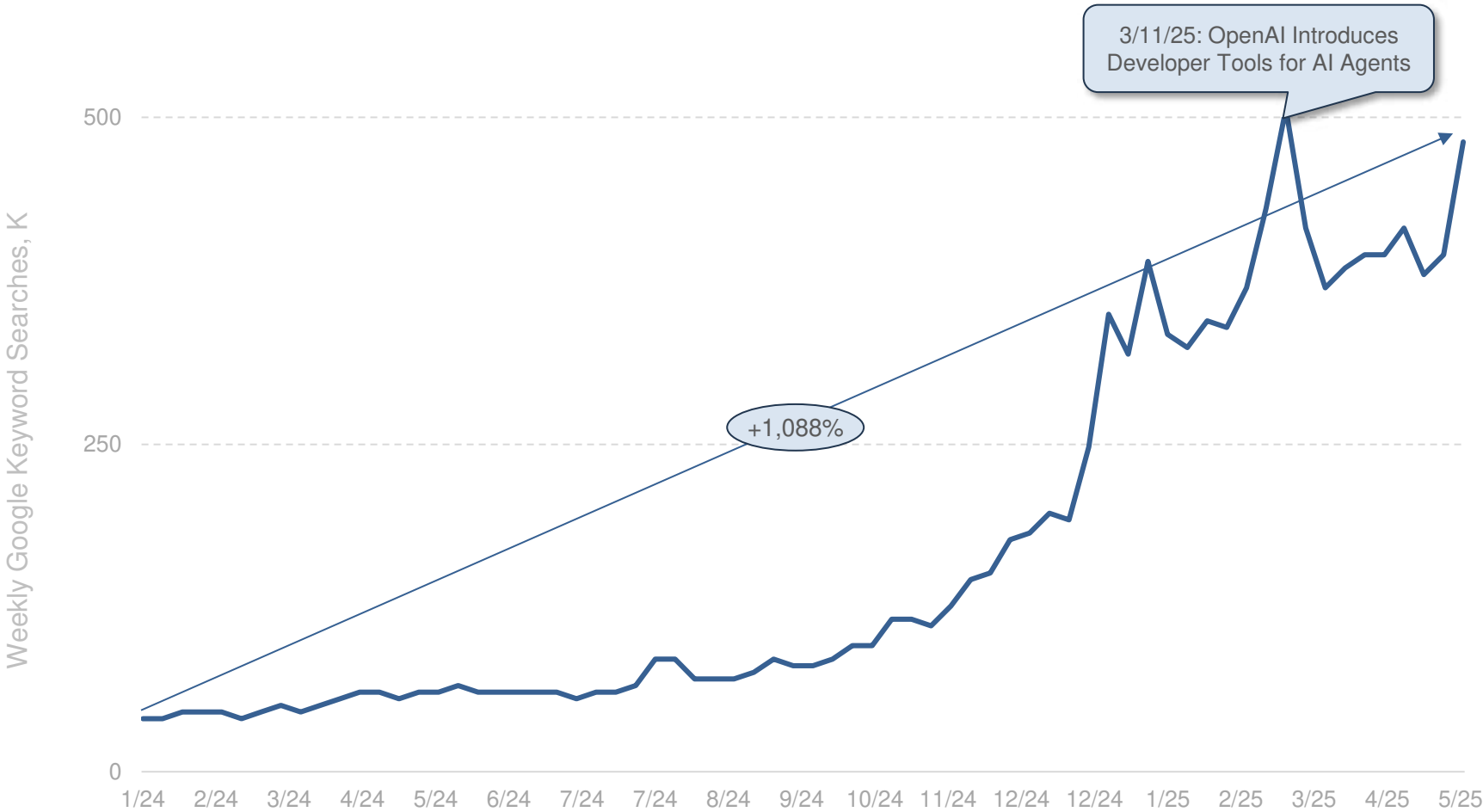
虽然我们还处于这些代理开发的早期阶段，但其影响才刚刚开始显现。AI 代理可能会重塑用户与数字系统交互的方式 – 从客户支持和引导到研究、日程安排和内部运营。

企业正在引领潮流；他们不仅在试验代理，还在部署它们，投资于框架并围绕自主执行构建生态系统。曾经的消息传递界面正在变成一个行动层。

AI Agent Interest (Google Searches) =  
+1,088% Over Sixteen Months

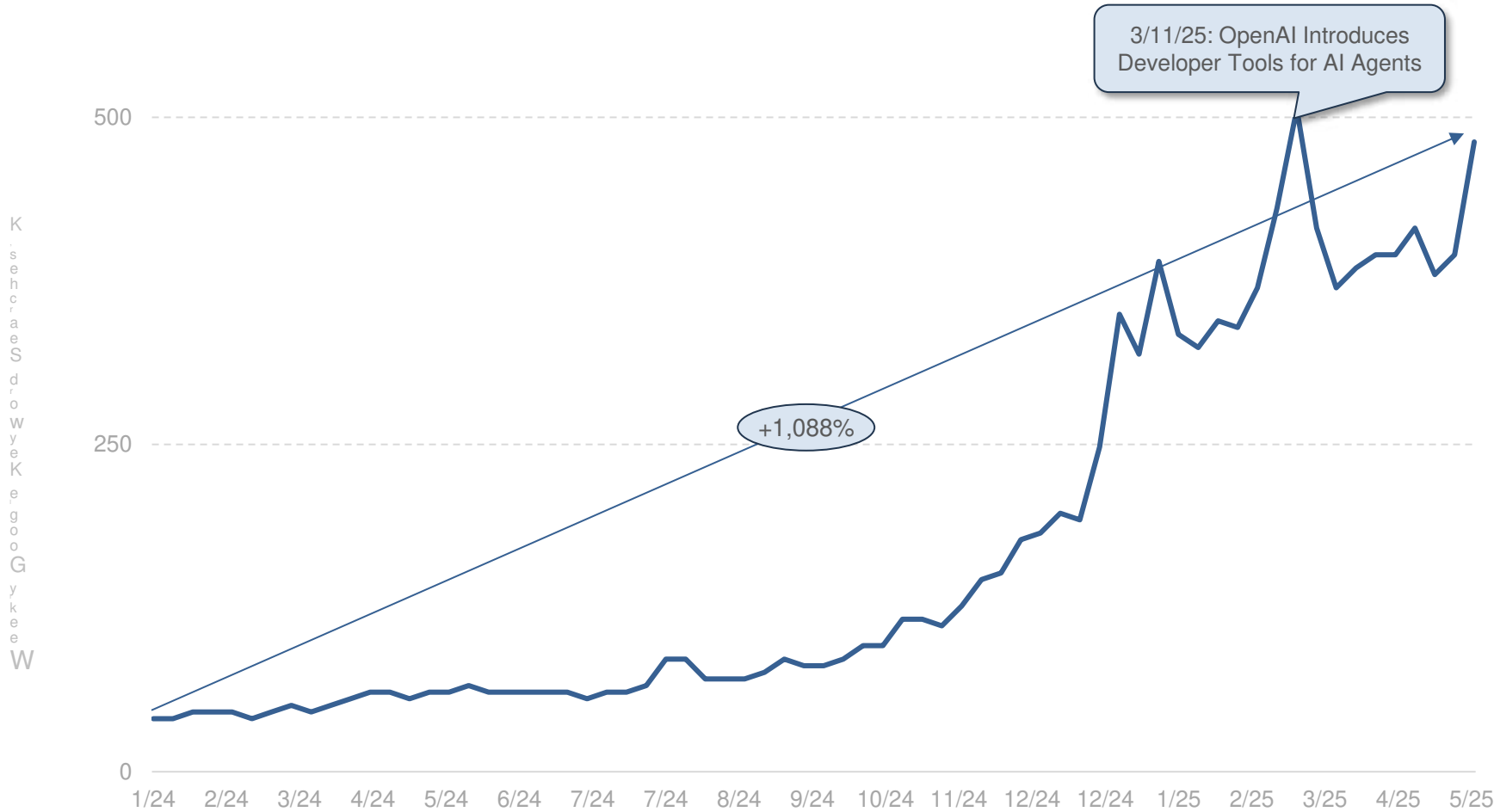
AI Agent Interest (Google Searches) =  
+1,088% Over Sixteen Months

Global Google Searches for ‘AI Agent’ (K) – 1/24-5/25, per Google Trends



Source: Google Trends via Glimpse (5/15/24), OpenAI (3/25)

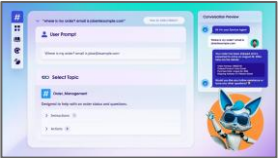



‘AI Agent’ (K) 的全球 Google 搜索 – 1/24-5/25, 根据 Google Trends



Source: Google Trends via Glimpse (5/15/24), OpenAI (3/25)



# AI Agent Deployments = AI Incumbent Product Launches Accelerating

AI Incumbent Agent Launches			
Agent Released		Select Capabilities	
	<b>Salesforce Agentforce</b> (10/24 = General Release)	<ul style="list-style-type: none"><li>Automated customer support</li><li>Case resolution</li><li>Lead qualification</li><li>Order tracking</li></ul>	
	<b>Anthropic Claude 3.5 Computer Use</b> (10/24 = Research Preview Release)	<ul style="list-style-type: none"><li>Control computer screen directly to perform tasks like pulling data from websites, making online purchases, etc.</li></ul>	
	<b>OpenAI Operator</b> (1/25 = Research Preview Release)	<ul style="list-style-type: none"><li>Control computer screen directly to perform tasks like pulling data from websites, making online purchases, etc.</li></ul>	
	<b>Amazon Nova Act</b> (3/25 = Research Preview Release)	<ul style="list-style-type: none"><li>Home automation</li><li>Information collection</li></ul>	<ul style="list-style-type: none"><li>Purchasing</li><li>Scheduling</li></ul>

Source: Salesforce (10/24), Salesforce Ben, Anthropic (10/24), OpenAI (1/25), Amazon (3/25)

# AI Agent Deployments = AI Incumbent Product Launches Accelerating

AI 行业巨头代理发布			
Agent 发布		选择功能	
	<b>Salesforce Agentforce</b> ( 10/24 = 正式发布 )	<ul style="list-style-type: none"><li>自动化客户支持案例解决</li><li></li></ul>	<ul style="list-style-type: none"><li>潜在客户资格认定</li><li>订单跟踪</li></ul>
	<b>Anthropic Claude 3.5 计算机使用</b> (10/24 = 研究预览版 )	<ul style="list-style-type: none"><li>直接控制电脑屏幕来执行诸如从网站提取数据、进行在线购买等任务。</li></ul>	
	<b>OpenAI Operator</b> ( 1/25 = 研究预览版 )	<ul style="list-style-type: none"><li>直接控制电脑屏幕来执行诸如从网站提取数据、进行在线购买等任务。</li></ul>	
	<b>Amazon Nova Act</b> (3/25 = 研究预览版 )	<ul style="list-style-type: none"><li>家庭自动化</li><li>信息收集</li></ul>	<ul style="list-style-type: none"><li>Purchasing</li><li>Scheduling</li></ul>

来源：Salesforce (10/24), Salesforce Ben, Anthropic (10/24), OpenAI (1/25), Amazon (3/25)



Next Frontier For AI =  
Artificial General Intelligence

AI = 的下一个前沿  
通用人工智能

**Artificial General Intelligence, or AGI**, refers to systems capable of performing the full range of human intellectual tasks – reasoning, planning, learning from small data samples, and generalizing knowledge across domains. Unlike current AI models, which excel within specific (albeit broad) boundaries, AGI would be able to operate fully flexibly across disciplines and solve unfamiliar problems without retraining. It represents a major milestone in AI development – one that builds on recent exponential gains in model scale, training data, and computational efficiency.

Timelines for AGI remain uncertain, but expert expectations have shifted forward meaningfully in recent years. Sam Altman, CEO of OpenAI, remarked in January 2025, *We are now confident we know how to build AGI as we have traditionally understood it*. This is a forecast, not a dictum, but it reflects how advances in model architecture, inference\* efficiency, and training scale are shortening the distance between research and frontier capability. The broader thread is clear: AI development is trending at unprecedented speed, and AGI is increasingly being viewed not as a hypothetical endpoint, but as a reachable threshold.

If / when achieved, AGI would redefine what software (and related hardware) can do. Rather than executing pre-programmed tasks, AGI systems would understand goals, generate plans, and self-correct in real time. They could drive research, engineering, education, and logistics workflows with little to no human oversight – handling ambiguity and novelty with general-purpose reasoning. These systems wouldn’t require extensive retraining to handle new problem domains – they would transfer learning and operate with context, much like human experts. Additionally, humanoid robots powered by AGI would have the power to reshape our physical environment and how we operate in it.

Still, the implications warrant a measured view. AGI is not a finish line, but a phase shift in capability – and how it reshapes institutions, labor, and decision-making will depend on the safeguards and deployment frameworks that accompany it. The productivity upside may be significant, but unevenly distributed.

The geopolitical, ethical, and economic implications may evolve gradually, not abruptly. As with earlier transitions – from industrial to digital to algorithmic – the full consequences will be shaped not just by what the technology can do, but by how society chooses to adopt and govern it.

*\*Inference = Fully-trained model generates predictions, answers, or content in response to user inputs. This phase is much faster and more efficient than training.*

通用人工智能，或 AGI，指的是能够执行全部人类智力任务的系统 – 推理、计划、从小数据样本中学习以及跨领域推广知识。与目前在特定（尽管范围广泛）边界内表现出色的 AI 模型不同，AGI 将能够跨学科完全灵活地运行，并在无需重新训练的情况下解决不熟悉的问题。它代表了 AI 发展的一个重要里程碑 – 它建立在模型规模、训练数据和计算效率的近期指数级增长的基础上。

AGI 的时间表仍然不确定，但近年来专家的预期已发生了有意义的转变。OpenAI 的首席执行官 Sam Altman 在 2025 年 1 月表示：我们现在确信我们知道如何构建我们传统上理解的 AGI。这是一种预测，而不是一种指示，但它反映了模型架构、推理 \* 效率和训练规模的进步如何缩短研究和前沿能力之间的距离。更广泛的主线是明确的：AI 的发展正以史无前例的速度发展，AGI 越来越被视为不是一个假设的终点，而是一个可以达到的阈值。

如果 / 当实现 AGI 时，AGI 将重新定义软件（以及相关硬件）的功能。AGI 系统将不再是执行预先编程的任务，而是理解目标、生成计划并实时进行自我纠正。它们可以在几乎不需要人工监督的情况下驱动研究、工程、教育和物流工作流程 – 通过通用 purpose reasoning. 处理模糊性和新颖性。这些系统不需要进行大量的再培训来处理新的问题领域 – 它们可以像人类专家一样进行迁移学习和上下文操作。此外，由 AGI 驱动的人类机器人将有能力重塑我们的物理环境以及我们在其中的运作方式。

尽管如此，其影响仍值得深思熟虑。AGI 不是终点线，而是能力上的一个相移 – 它如何重塑机构、劳动力和决策将取决于随之而来的保障措施和部署框架。生产力的提升可能非常显著，但分布不均。

地缘政治、伦理和经济影响可能会逐渐演变，而不是突然发生。与之前的转型一样 – 从工业到数字再到算法 – 最终的结果不仅取决于技术能够做什么，还取决于社会如何选择采用和管理它。

\* 推理 = 完全训练的模型生成预测、答案或内容，以响应用户输入。此阶段比训练快得多且效率更高

*g.*

*AI User + Usage + **CapEx Growth** =*  
*Unprecedented*

AI 用户 + 使用量 + **资本支出增长** =  
前所未有

To understand where technology CapEx is heading, it helps to look at where it’s been. Over the past two decades, tech CapEx has flexed upward at points through data’s long arc – first toward storage / access, then toward distribution / scale, and now toward computation / intelligence.

The earliest wave saw CapEx pouring into building internet infrastructure – massive server farms, undersea cables, and early data centers that enabled Amazon, Microsoft, Google and others to lay the foundation for cloud computing. That was the first phase: store it, organize it, serve it.

The second wave – still unfolding – has been about supercharging compute for data-heavy AI workloads, a natural evolution of cloud computing. Hyperscaler\* CapEx budgets now tilt increasingly toward specialized chips (GPUs, TPUs, AI accelerators…), liquid cooling, and frontier data center design.

In 2019, AI was a research feature; by 2023, it was a capital expenditure line item. Microsoft Vice Chair and President Brad Smith put it well in a 4/25 blog post: *Like electricity and other general-purpose technologies in the past, AI and cloud datacenters represent the next stage of industrialization.*

The world's biggest tech companies are spending tens of billions annually – not just to gather data, but to learn from it, reason with it and monetize it in real time. It’s still about data – but now, the advantage goes to those who can train on it fastest, personalize it deepest, and deploy it widest.

*\*Hyperscalers (large data center operators) are Amazon Web Services (AWS), Microsoft Azure, Google Cloud Platform (GCP), Alibaba Cloud, Oracle Cloud Infrastructure (OCI), IBM Cloud & Tencent Cloud.*

要了解技术资本支出的发展方向，了解其过去的发展历程会有所帮助。在过去的二十年中，技术资本支出在数据的长弧线上不断向上调整 —— 首先是存储 / 访问，然后是分发 / 规模，现在是计算 / 智能。

最早的一波浪潮是将资本支出投入到互联网基础设施的建设中 —— 大型服务器集群、海底电缆和早期数据中心，这些设施使 Amazon 、 Microsoft 、 Google 等公司能够为云计算奠定基础。那是第一阶段：存储、组织、服务。

第二波浪潮 —— 仍在展开 —— 一直是关于为数据密集型 AI 工作负载提供超级计算能力，这是云计算的自然演变。超大规模企业 \* 的资本支出预算现在越来越倾向于专用芯片（ GPU 、 TPU 、 AI 加速器 …… ）、液体冷却和前沿数据中心设计。

2019 年，AI 是一项研究功能；到 2023 年，它已成为一项资本支出项目。微软副董事长兼总裁 Brad Smith 在 4 月 25 日的博文中很好地阐述了这一点：就像过去的电力和其他通用技术一样，AI 和云数据中心代表着工业化的下一个阶段。

全球最大的科技公司每年花费数百亿美元 —— 不仅仅是为了收集数据，而是为了从中学习、推理并在实时中将其货币化。这仍然与数据有关 —— 但现在，优势属于那些能够以最快速度进行训练、最深入地进行个性化定制以及最广泛地进行部署的公司。

\* 超大规模企业（大型数据中心运营商）包括 Amazon Web Services (AWS)、 Microsoft Azure 、 Google Cloud Platform (GCP)、 Alibaba Cloud 、 Oracle Cloud Infrastructure (OCI)、 IBM Cloud 和 Tencent Cloud。

CapEx Spend – Big Technology Companies =

On Rise for Years as  
Data Use + Storage Exploded

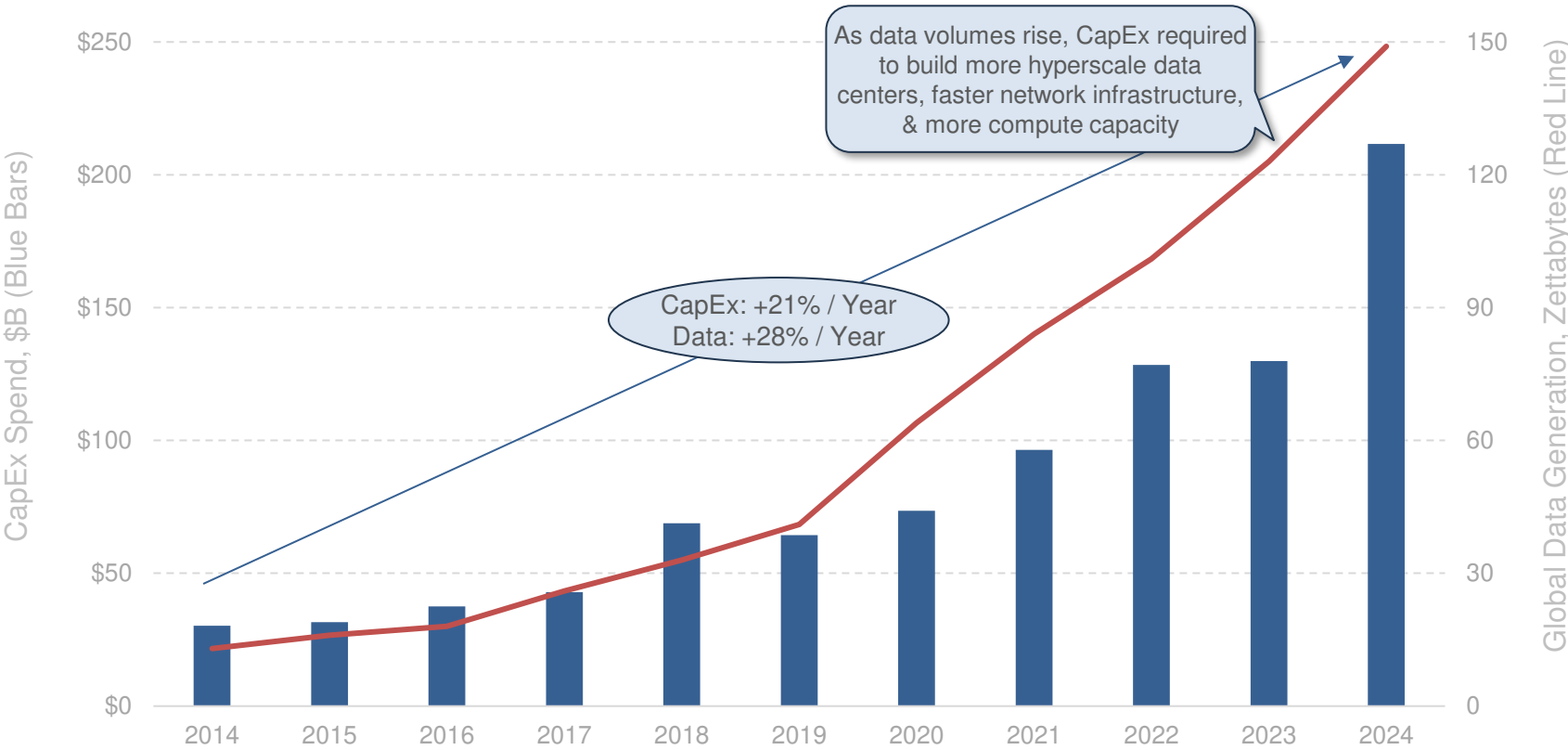
资本支出 – 大型科技公司 =

多年来随着数据使用量 + 存  
储的爆炸式增长而上升

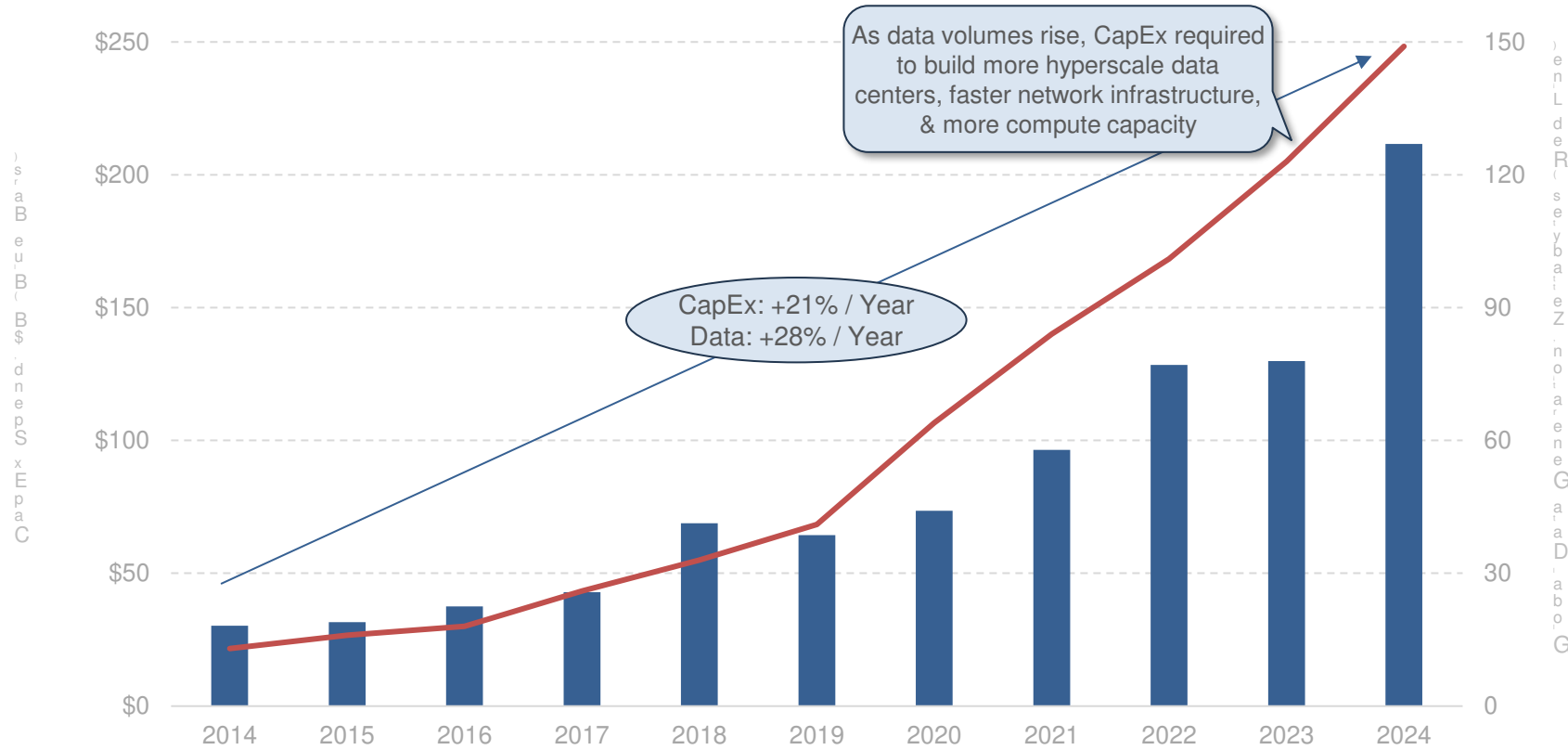
# CapEx Spend @ Big Six\* Tech Companies (USA) = +21% Annual Growth Over Ten Years

# 美国六大科技公司 \* 的资本支出 =+21% 近十年年均增长率

Big Six\* USA Public Technology Company CapEx Spend (\$B) vs. Global Data Generation (Zettabytes) – 2014-2024, per Capital IQ & Hinrich Foundation



美国六大 \* 上市科技公司资本支出（十亿美元）与全球数据生成量（泽字节） – 2014-2024 年，数据来源：Capital IQ 和 Hinrich Foundation



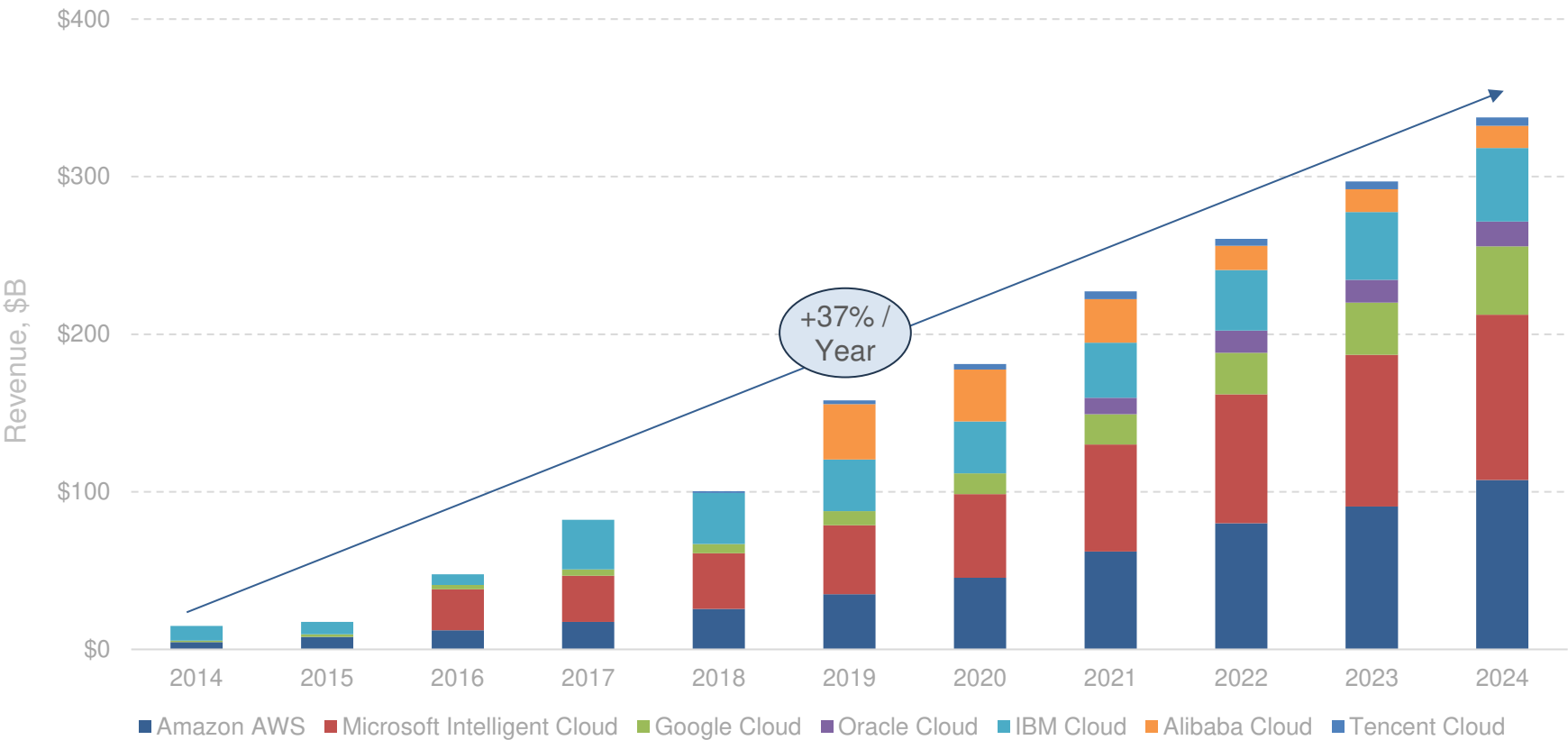
\*Note: Big Six USA technology companies include Apple, Nvidia, Microsoft, Alphabet / Google, Amazon, & Meta Platforms / Facebook. Only AWS CapEx & revenue shown for Amazon (i.e. excludes Amazon retail CapEx). AWS CapEx estimated per Morgan Stanley – equals AWS net additions to property & equipment less finance leases and obligations. Global data generation figures for 2024 are estimates. Source: Capital IQ (3/25), Hinrich Foundation (3/25)

\* 注：美国六大科技公司包括苹果、英伟达、微软、Alphabet / 谷歌、亚马逊和 Meta Platforms / Facebook。亚马逊仅显示 AWS 资本支出和收入（即不包括亚马逊零售资本支出）。AWS 资本支出根据摩根士丹利估算 – 等于 AWS 净增加的财产和设备减去融资租赁和义务。2024 年的全球数据生成量数据为估计值。来源：Capital IQ (3/25), Hinrich Foundation (3/25)

CapEx Spend for Tech Hyperscalers = Mirrored by...  
+37% Annual Cloud Revenue Growth Over Ten Years

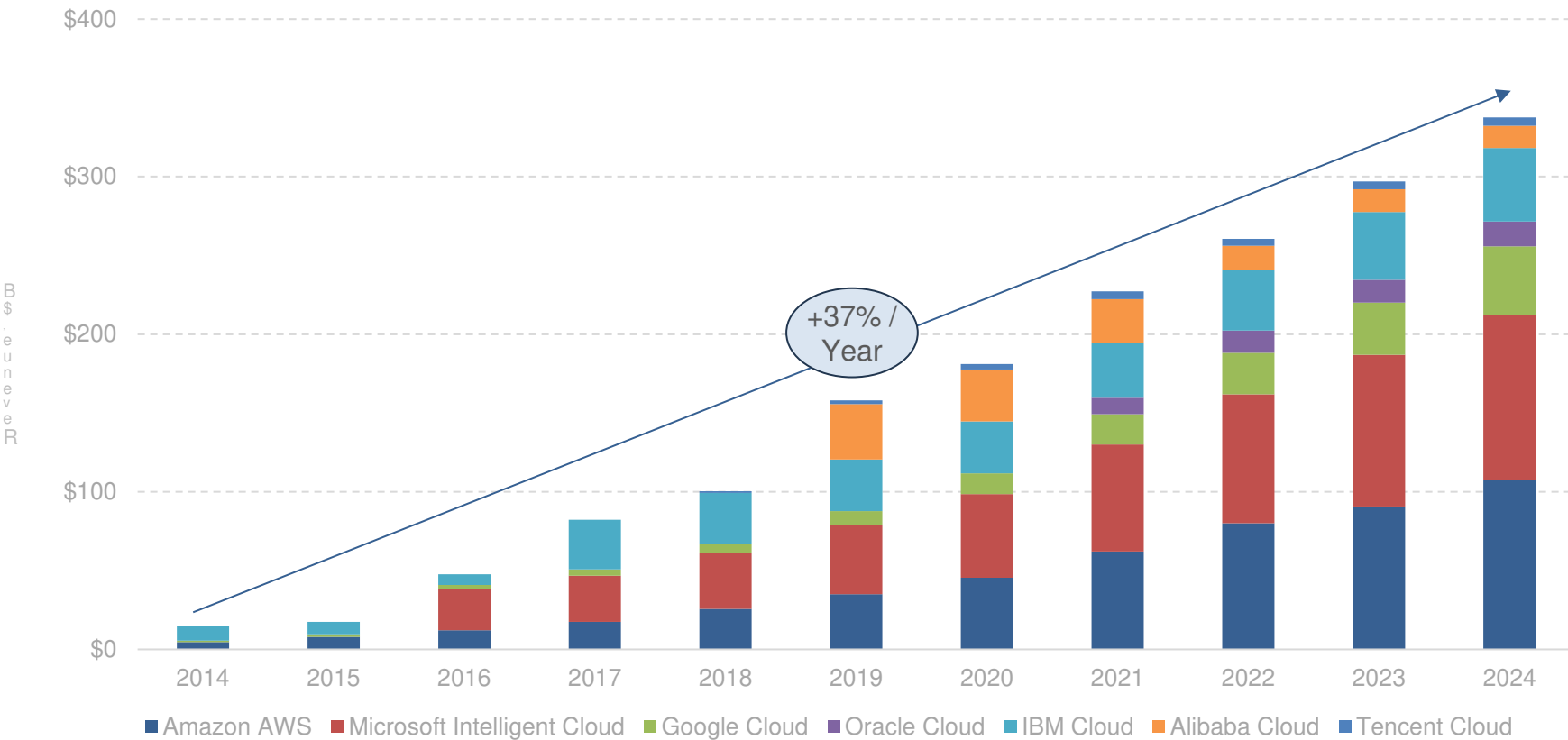
科技巨头 = 的资本支出反映了 ……+37% 的十年年均云收入  
增长

Global Hyperscaler Cloud Revenue (\$B) – 2014-2024,  
per Company Disclosures & Morgan Stanley Estimates



Note: Companies do not report “hyperscaler cloud revenue” on like-for-like basis so data represents best estimates and may not align between companies. Oracle Cloud revenue includes Cloud Services & License Support, as well as Cloud License & On-Premise License. IBM Cloud includes all ‘Infrastructure’ line items due to reporting standards. Alibaba & Tencent Cloud revenues estimated per Morgan Stanley. Source: Company disclosures, Morgan Stanley (as of 4/25)

全球超大规模企业云收入（十亿美元）–2014-2024 年，根据  
公司披露和摩根士丹利估计



注：各公司并未在同等基础上报告“超大规模企业云收入”，因此数据为最佳估算值，可能与各公司的数据不一致。Oracle Cloud 收入包括云服务和许可支持以及云许可和本地许可。由于报告标准，IBM Cloud 包括所有“基础设施”行项目。阿里巴巴和腾讯云收入由摩根士丹利估算。资料来源：公司披露，摩根士丹利（截至 4/25）

CapEx Spend – Big Technology Companies =

Inflected With AI's Rise

CapEx 支出 – 大型科技公司 =

Inflected With AI's Rise

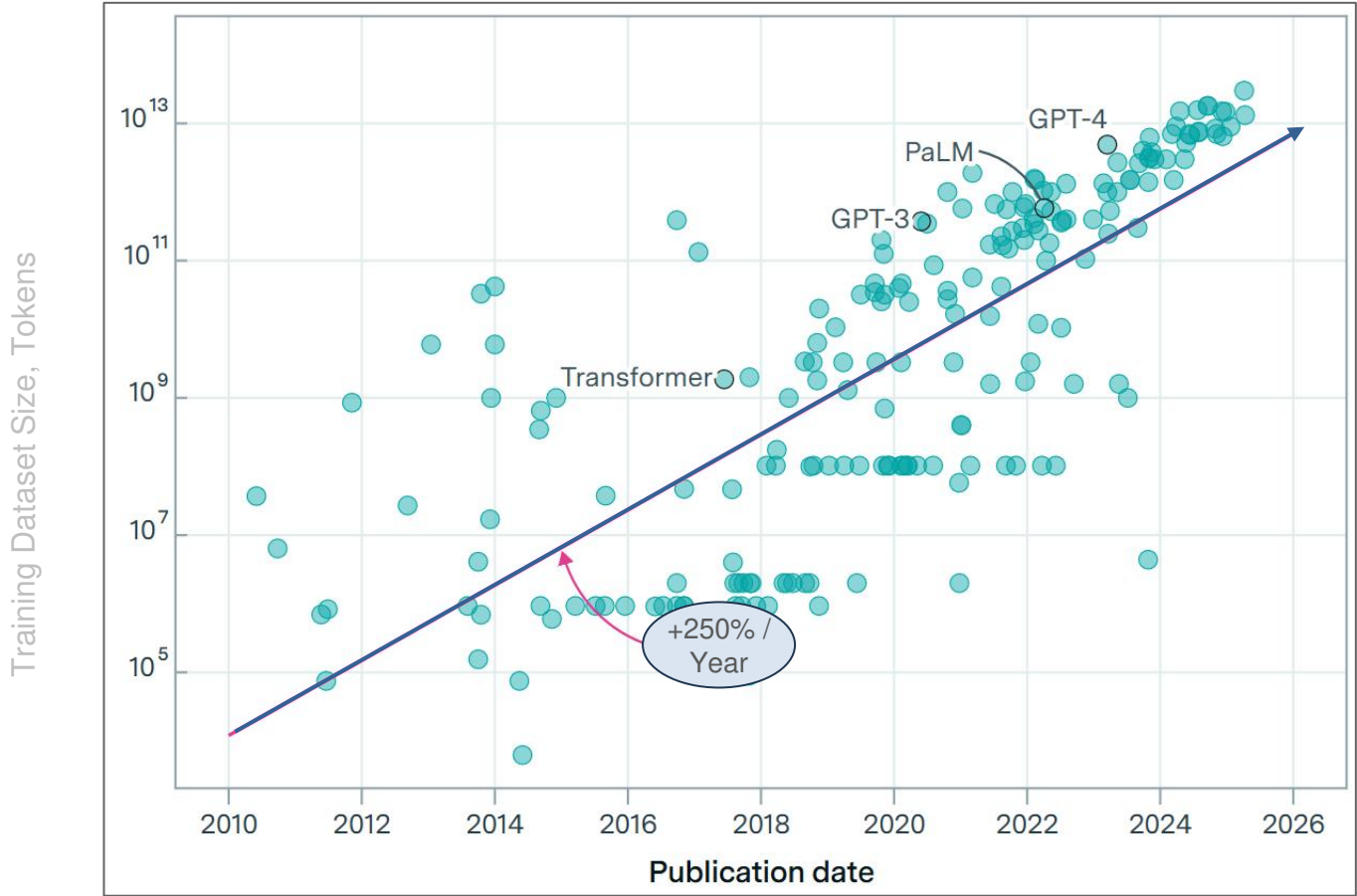


AI Model Training Dataset Size = 250% Annual Growth Over Fifteen Years, per Epoch AI

AI 模型训练数据集大小 = 每年增长 250%，持续 15 年，每次 Epoch AI

AI Model Training Dataset Size (Tokens) by Model Release Year – 6/10-5/25, per Epoch AI

按模型发布年份划分的 AI 模型训练数据集大小（Token） – 6/10-5/25，每次 Epoch AI

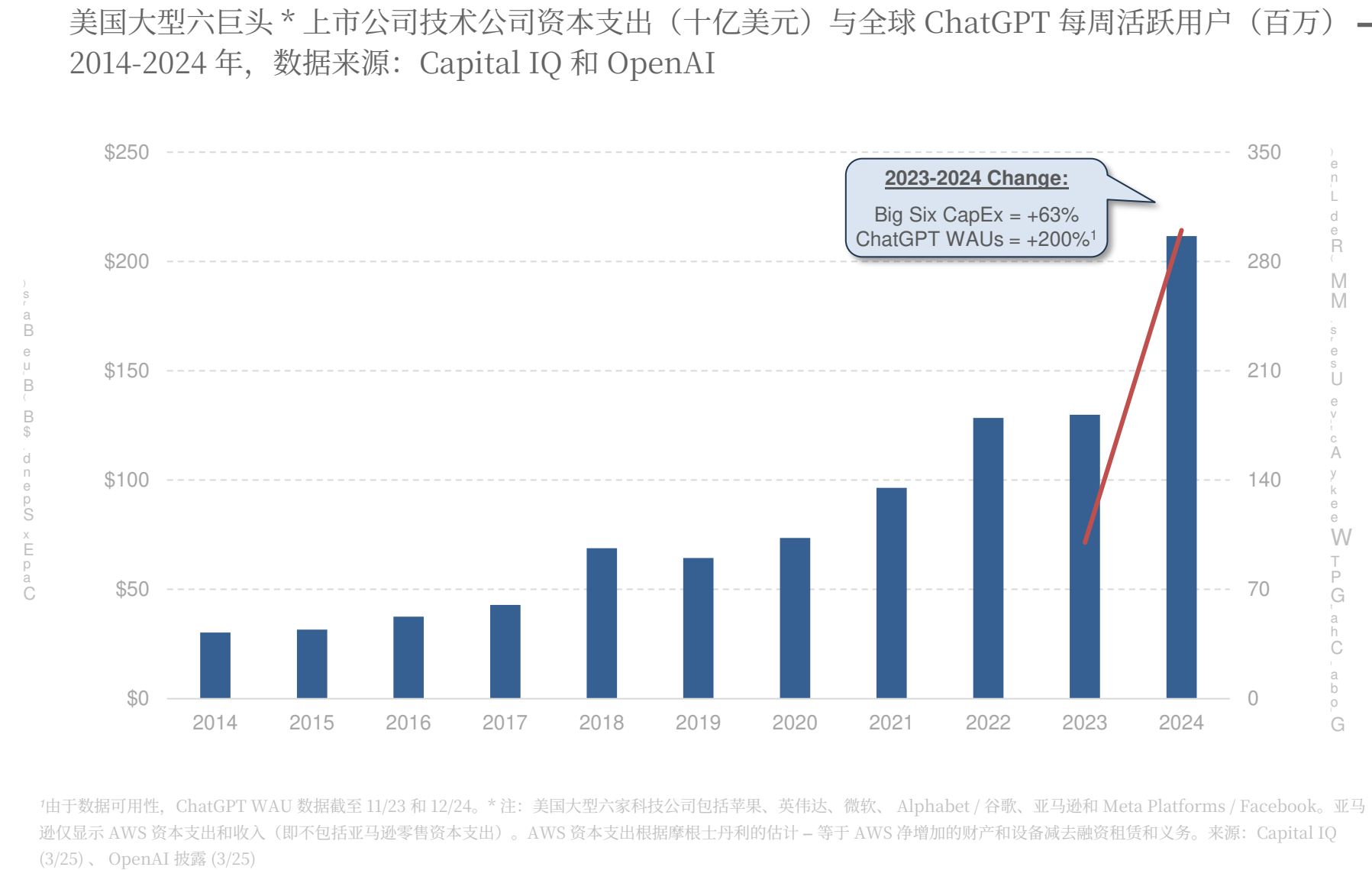
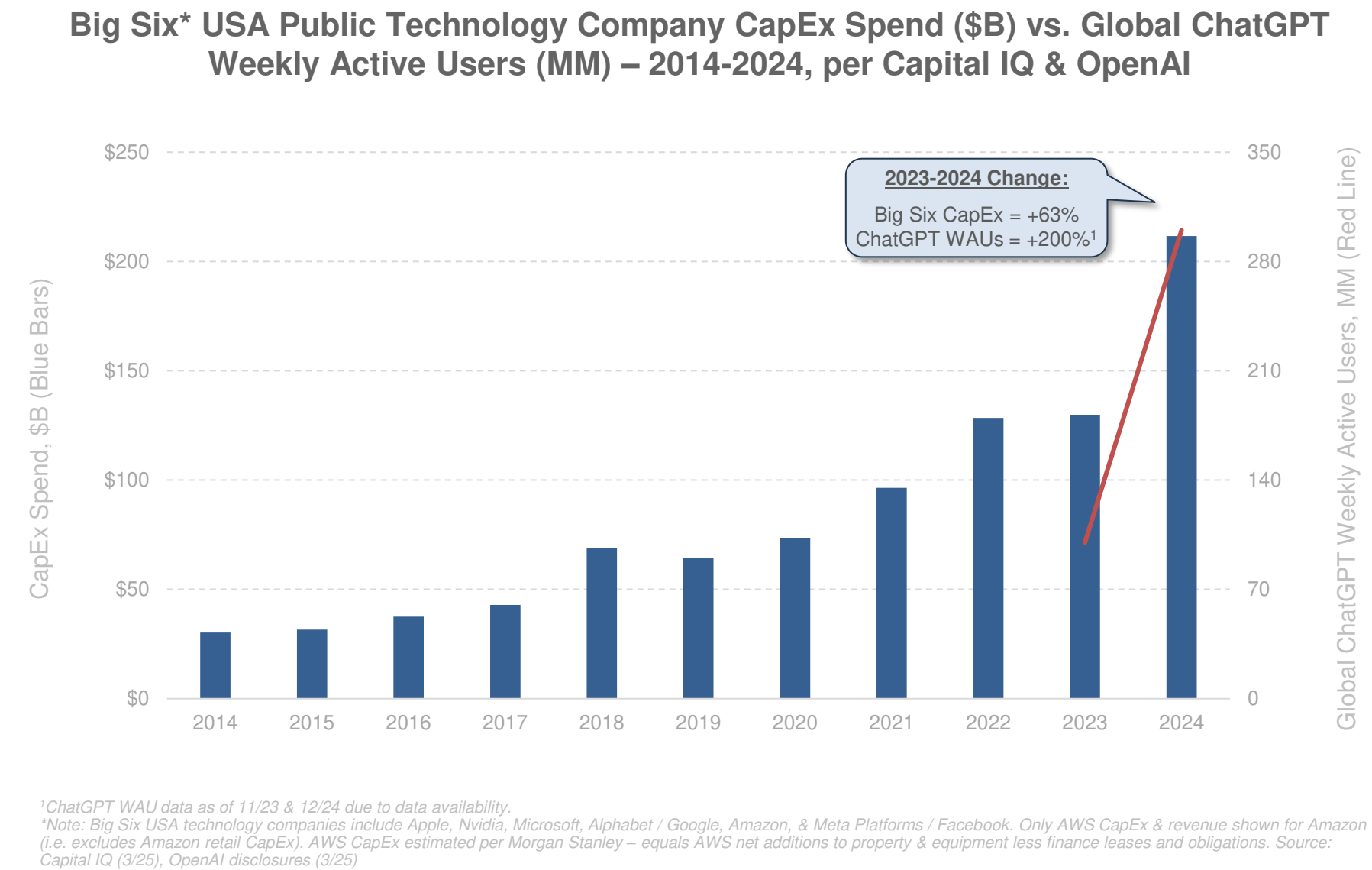


Note: In AI language models, tokens represent basic units of text (e.g., words or sub-words) used during training. Training dataset sizes are often measured in total tokens processed. A larger token count typically reflects more diverse and extensive training data, which can lead to improved model performance – up to a point – before reaching diminishing returns. Source: Epoch AI (5/25)

注意：在 AI 语言模型中，token 代表在训练期间使用的文本基本单位（例如，单词或子词）。训练数据集的大小通常以处理的 token 总数来衡量。较大的 token 计数通常反映更多样化和广泛的训练数据，这可以提高模型性能 – 达到一定程度 – 之后才会达到收益递减。来源：Epoch AI (5/25)

CapEx Spend @ Big Six\* Tech Companies =  
+63% Y/Y & Accelerated...

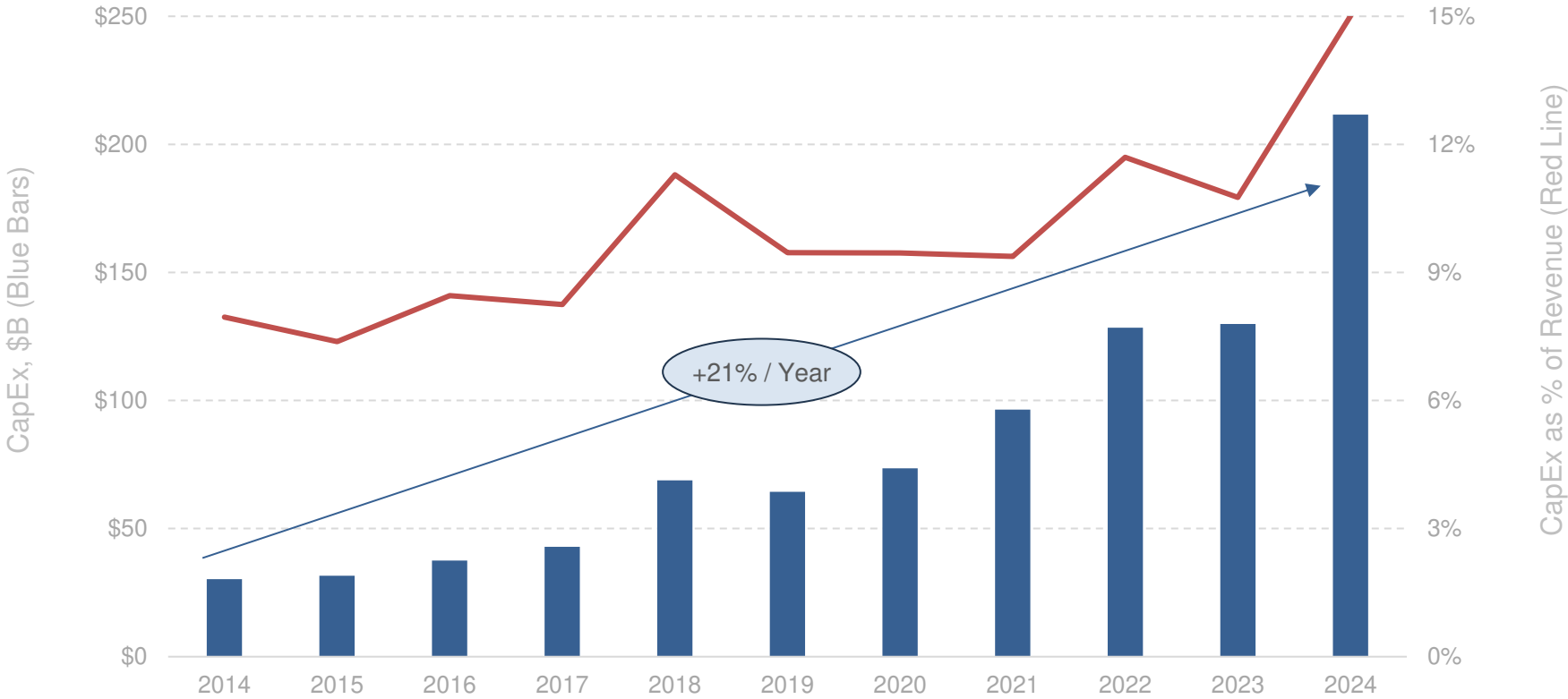
大型六巨头 \* 科技公司资本支出 =+63% 同比增长  
& 加速 ...



...CapEx Spend @ Big Six\* Tech Companies = 15% of Revenue & Accelerated vs. 8% Ten Years Ago

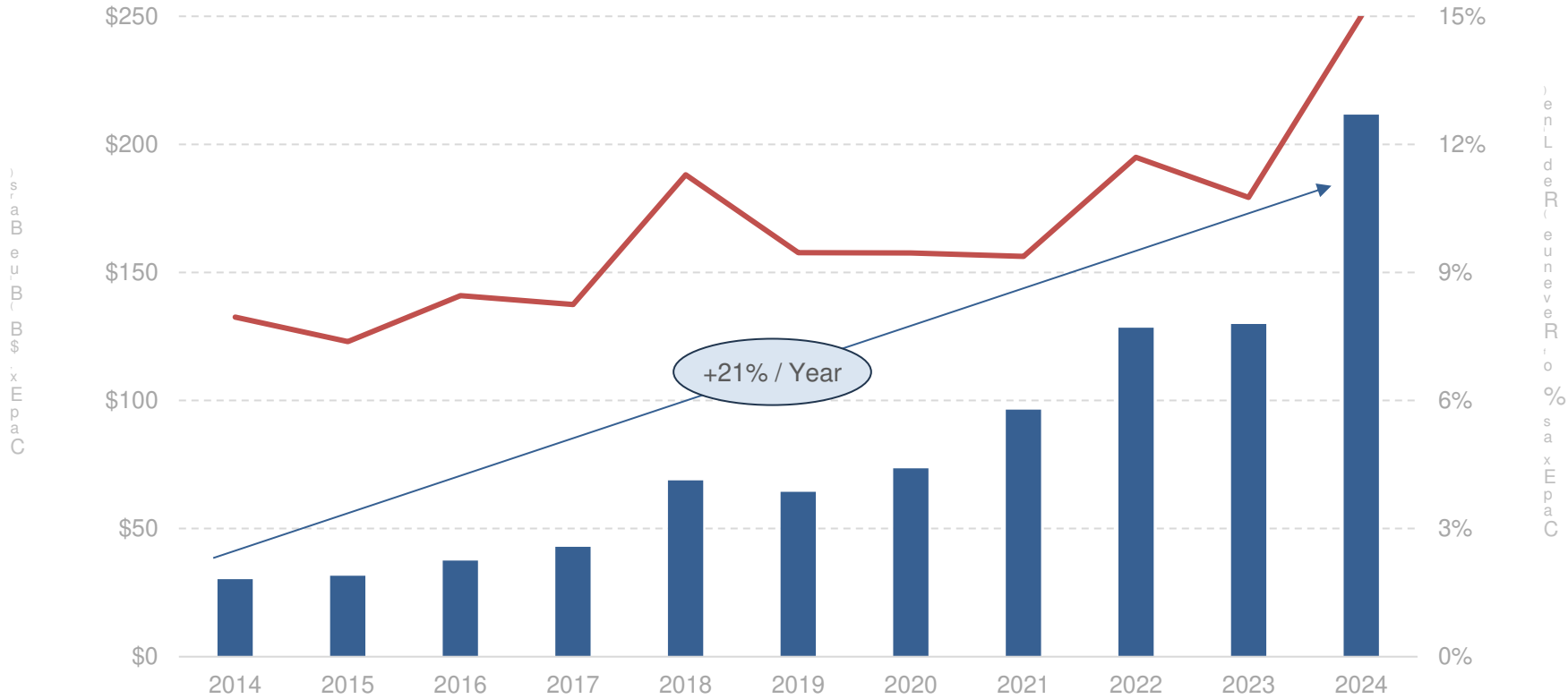
... 大型科技公司（Big Six\*）的资本支出 = 占收入的 15%，且增速加快，而十年前为 8%

Big Six\* USA Public Technology Company – CapEx Spend (\$B) vs. % of Revenue – 2014-2024, per Capital IQ & Morgan Stanley



\*Note: Big Six USA technology companies include Apple, Nvidia, Microsoft, Alphabet / Google, Amazon, & Meta Platforms / Facebook. Only AWS CapEx & revenue shown for Amazon (i.e. excludes Amazon retail CapEx). AWS CapEx estimated per Morgan Stanley – equals AWS net additions to property & equipment less finance leases and obligations. Source: Capital IQ (3/25), Morgan Stanley (5/25)

美国六大 \* 上市科技公司 – 资本支出（十亿美元）与收入百分比 –2014-2024 年，数据来源：Capital IQ 和摩根士丹利



\* 注：美国六大科技公司包括苹果、英伟达、微软、Alphabet / 谷歌、亚马逊和 Meta Platforms / Facebook。亚马逊仅显示 AWS 的资本支出和收入（即不包括亚马逊零售的资本支出）。AWS 的资本支出根据摩根士丹利估算 – 等于 AWS 的财产和设备净增加额减去融资租赁和义务。资料来源：Capital IQ (3/25), Morgan Stanley (5/25)

CapEx Spend @ Amazon AWS =

Cloud vs. AI Patterns

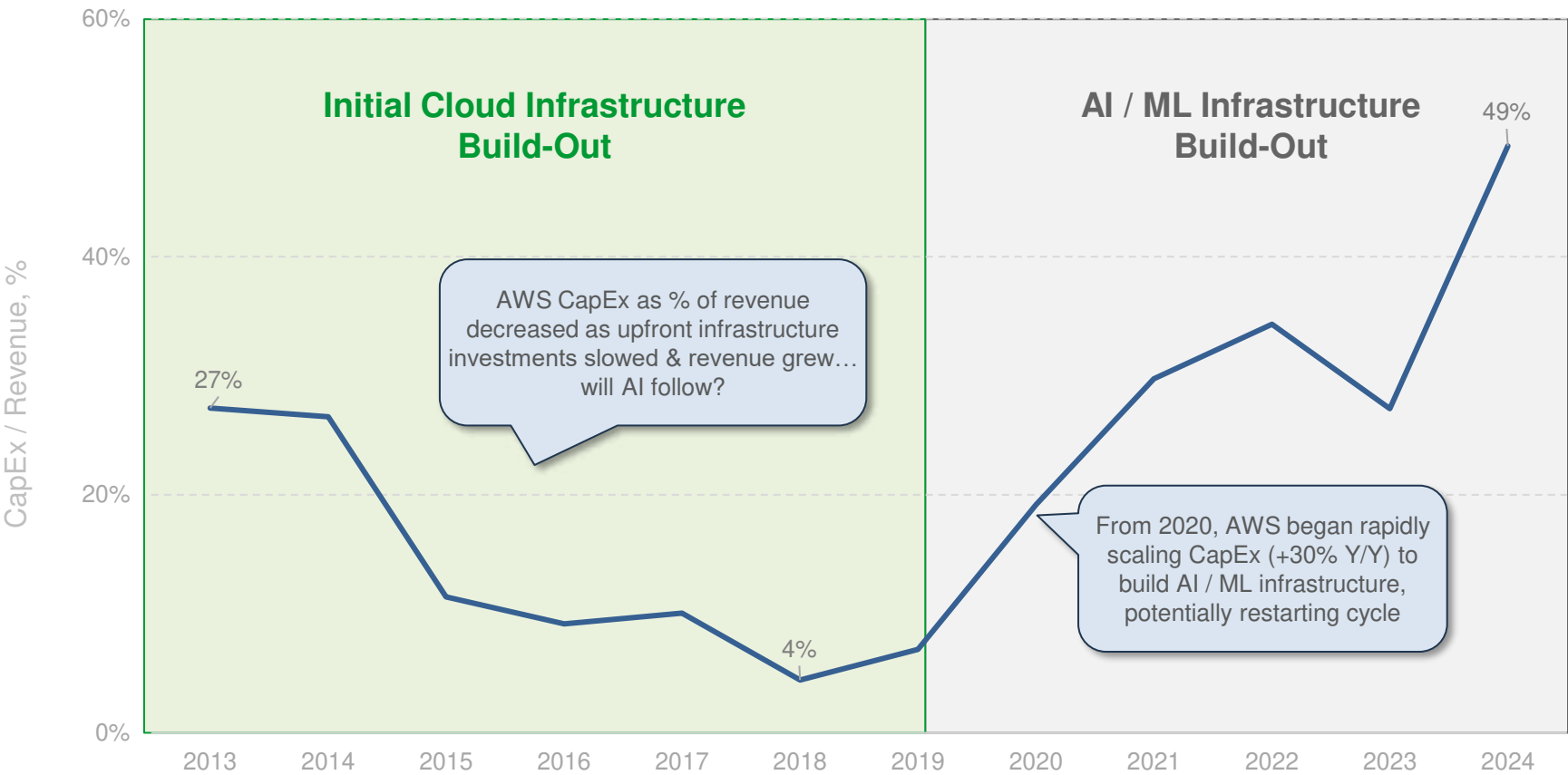
CapEx Spend @ Amazon AWS =

云与人工智能模式

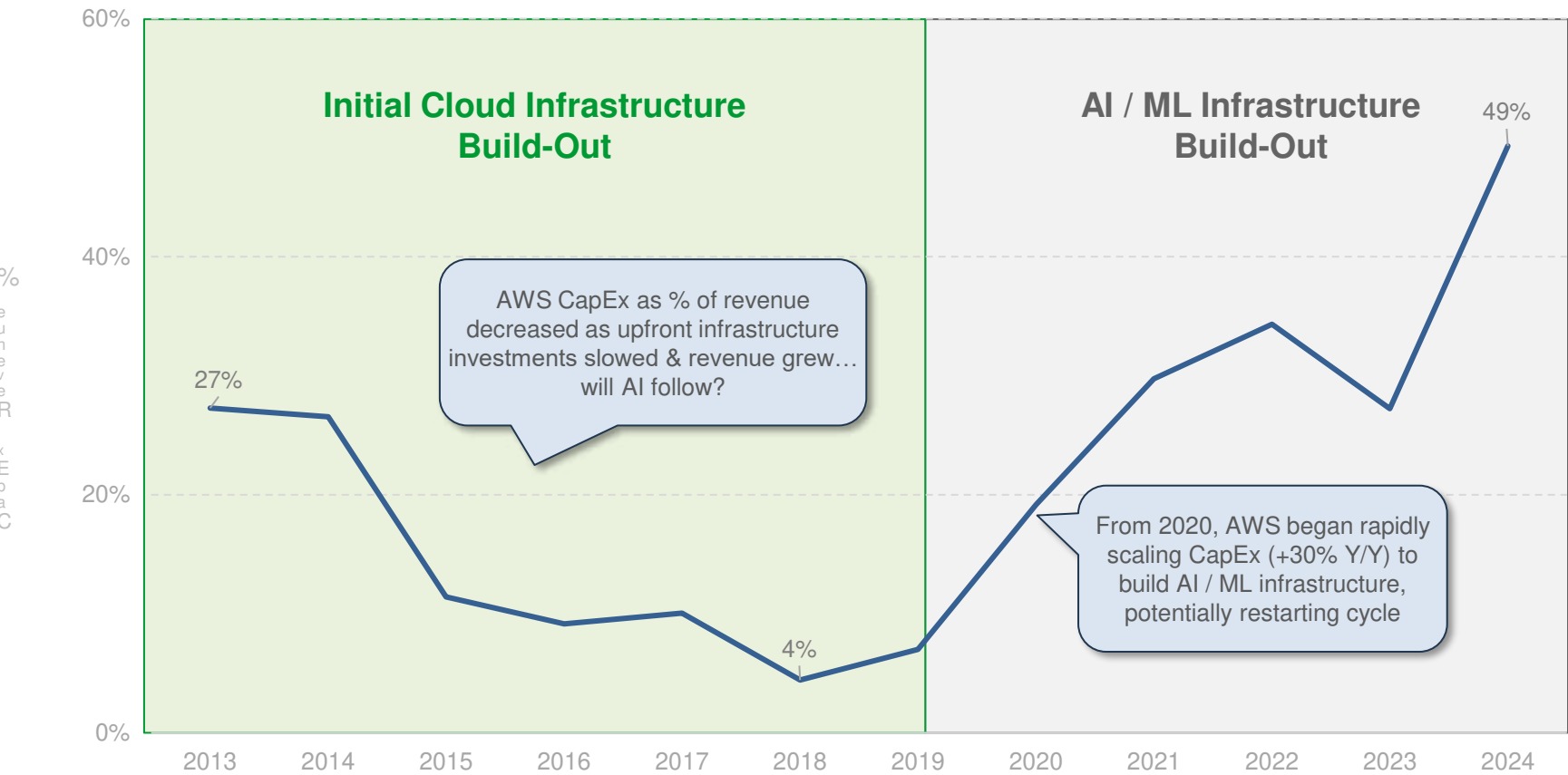
CapEx as % of Revenue (AWS as Proxy) – AI vs. Cloud Buildouts = 49% (2024) vs. 4% (2018) vs. 27% (2013), per Morgan Stanley

资本支出占收入的百分比（以 AWS 为代表）–AI 与云基础设施建设对比 =49% （2024 年） vs. 4% （2018 年） vs. 27% （2013 年）， 数据来源：摩根士丹利

Amazon AWS CapEx as % of Revenue – 2013-2024, Estimated per Morgan Stanley



Amazon AWS CapEx as % of Revenue – 2013-2024, Estimated per Morgan Stanley



Note: Figures shown represent AWS only. AWS CapEx estimated per Morgan Stanley – equals AWS net additions to property & equipment less finance leases and obligations. Source: Amazon, Morgan Stanley (5/25)

注意：显示的数字仅代表 AWS。根据摩根士丹利的估算，AWS 的资本支出 – 等于 AWS 的财产和设备净增加额减去融资租赁和义务。来源：亚马逊，摩根士丹利（5/25）

Tech CapEx Spend Partial Instigator =  
Material Improvements in GPU Performance

技术资本支出部分推动因素 =  
GPU 性能的重大改进

# NVIDIA GPU Performance = +225x Over Eight Years

# NVIDIA GPU 性能在八年内提升 =+225 倍

Performance of NVIDIA GPU Series Over Time – 2016-2024, per NVIDIA

\$1B Data Center Comparison GPT-MoE Inference Workload <sup>1</sup>						
	Pascal	Volta	Ampere	Hopper	Blackwell	
	2016	2018	2020	2022	2024	
Number of GPUs	46K	43K	28K	16K	11K	+225x
Factory AI FLOPS	1EF	5EF	17EF	63EF	220EF	
Annual Inference Tokens	50B	1T	5T	58T	1,375T	+30,000x
Annual Token Revenue	\$240K	\$3M	\$24M	\$300M	\$7B	
DC Power	37MW	34MW	25MW	19MW	21MW	+50,000x
Token Per MW-Year	1.3B	2.9B	200B	3T	65T	

For a Theoretical \$1B-Scale Data Center...

...Performance +225x over eight years while requiring 4x fewer GPUs...

...Inference token capacity +27,500x over eight years, implying +30,000x higher theoretical token revenue...

...Data center power use down 43% over eight years, leading to +50,000x greater per-unit energy efficiency

NVIDIA GPU 系列随时间推移的性能 – 2016-2024，数据来源：NVIDIA

10 亿美元数据中心对比 GPT-MoE 推理工作负载 <sup>1</sup>						
	Pascal	Volta	Ampere	Hopper	Blackwell	
	2016	2018	2020	2022	2024	
Number of GPUs	46K	43K	28K	16K	11K	+225x
Factory AI FLOPS	1EF	5EF	17EF	63EF	220EF	
Annual Inference Tokens	50B	1T	5T	58T	1,375T	+30,000x
Annual Token Revenue	\$240K	\$3M	\$24M	\$300M	\$7B	
DC Power	37MW	34MW	25MW	19MW	21MW	+50,000x
Token Per MW-Year	1.3B	2.9B	200B	3T	65T	

For a Theoretical \$1B-Scale Data Center...

...Performance +225x over eight years while requiring 4x fewer GPUs...

...Inference token capacity +27,500x over eight years, implying +30,000x higher theoretical token revenue...

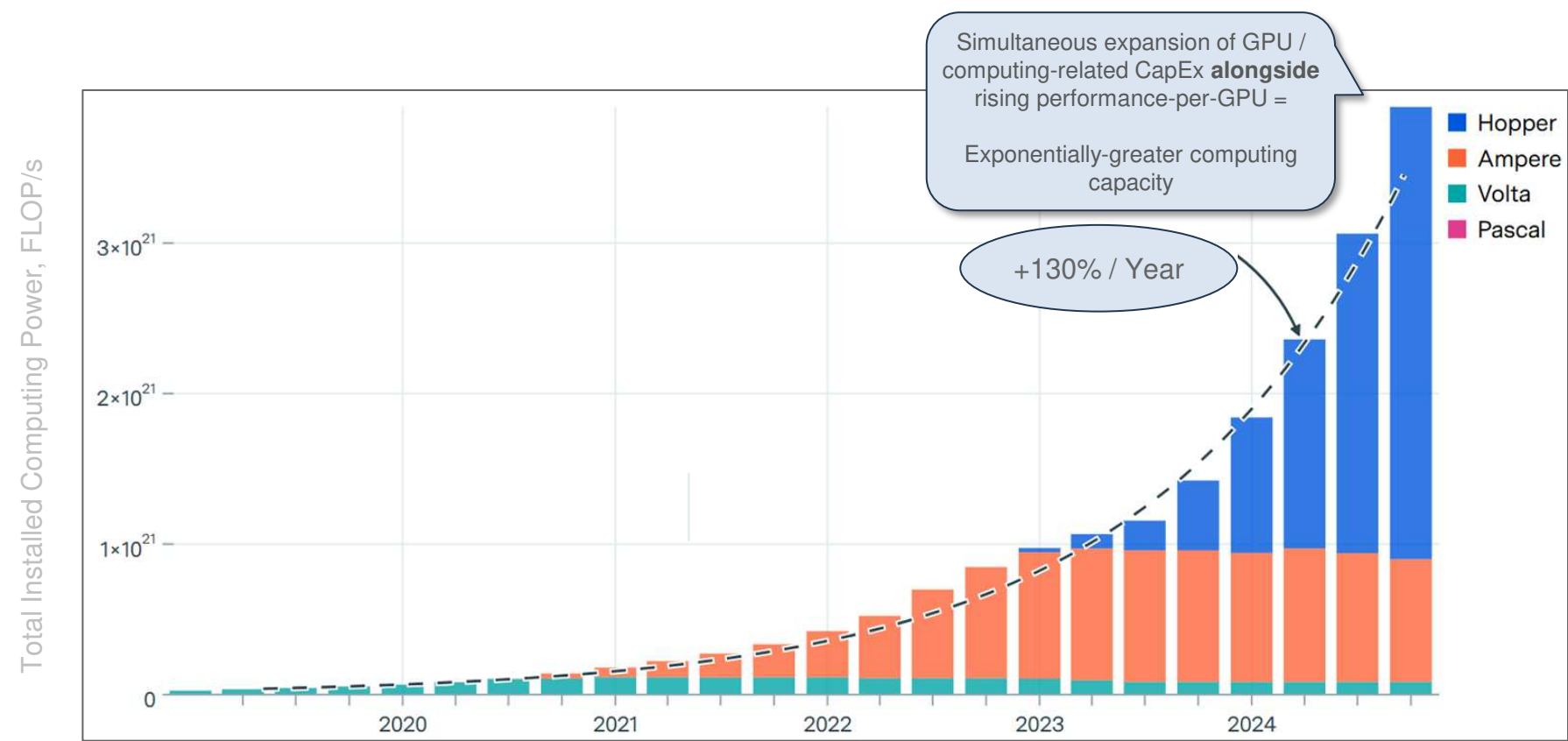
...Data center power use down 43% over eight years, leading to +50,000x greater per-unit energy efficiency

<sup>1</sup> GPT-MoE Inference Workload = A type of workload where a GPT-style model with a Mixture-of-Experts (MoE) architecture is used for inference (i.e., making predictions).  
Note: Annual token revenue assumes a flat per-token cost. Source: NVIDIA (5/25)

<sup>1</sup> GPT-MoE 推理工作负载 = 一种工作负载类型，其中具有混合专家 (MoE) 架构的 GPT 风格模型用于推理（即，进行预测）。注意：年度代币收入假设每个代币的成本不变。来源：NVIDIA (5/25)

# NVIDIA Installed GPU Computing Power = 100x+ Growth Over ~Six Years

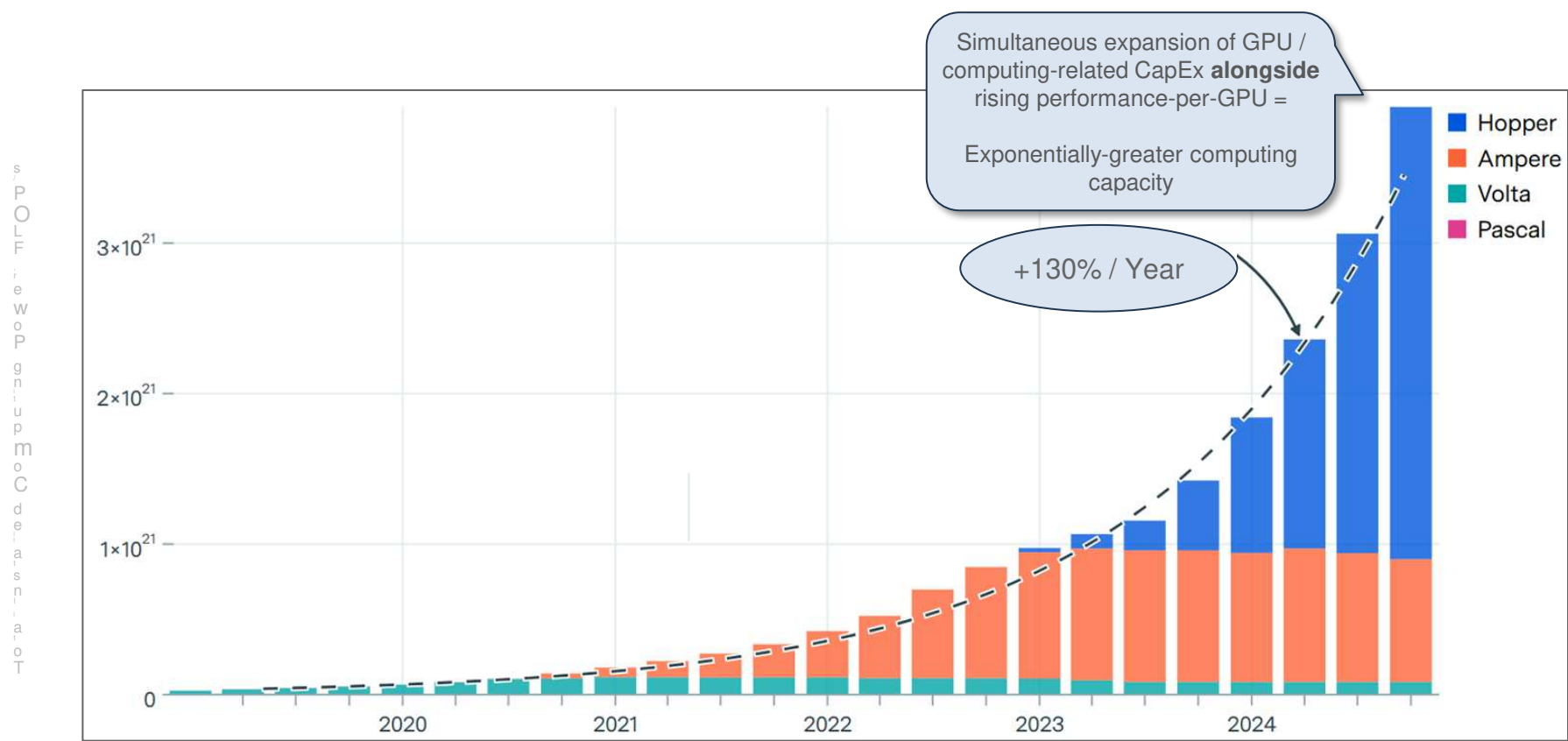
Global Stock of NVIDIA GPU Computing Power (FLOP/s) – Q1:19-Q4:24, per Epoch AI



Note: Analysis does not include TPUs or other specialized AI accelerators, for which less data is available. TPUs may provide comparable total computing power to NVIDIA chips. Source: Epoch AI (2/25)

# NVIDIA Installed GPU Computing Power = 100x+ Growth Over ~Six Years

Global Stock of NVIDIA GPU Computing Power (FLOP/s) – Q1:19-Q4:24, per Epoch AI



注意：分析不包括 TPU 或其他专用 AI 加速器，因为这些加速器的数据较少。TPU 可能提供与 NVIDIA 芯片相当的总计算能力。来源：Epoch AI (2/25)



Tech CapEx Spend Beneficiary =

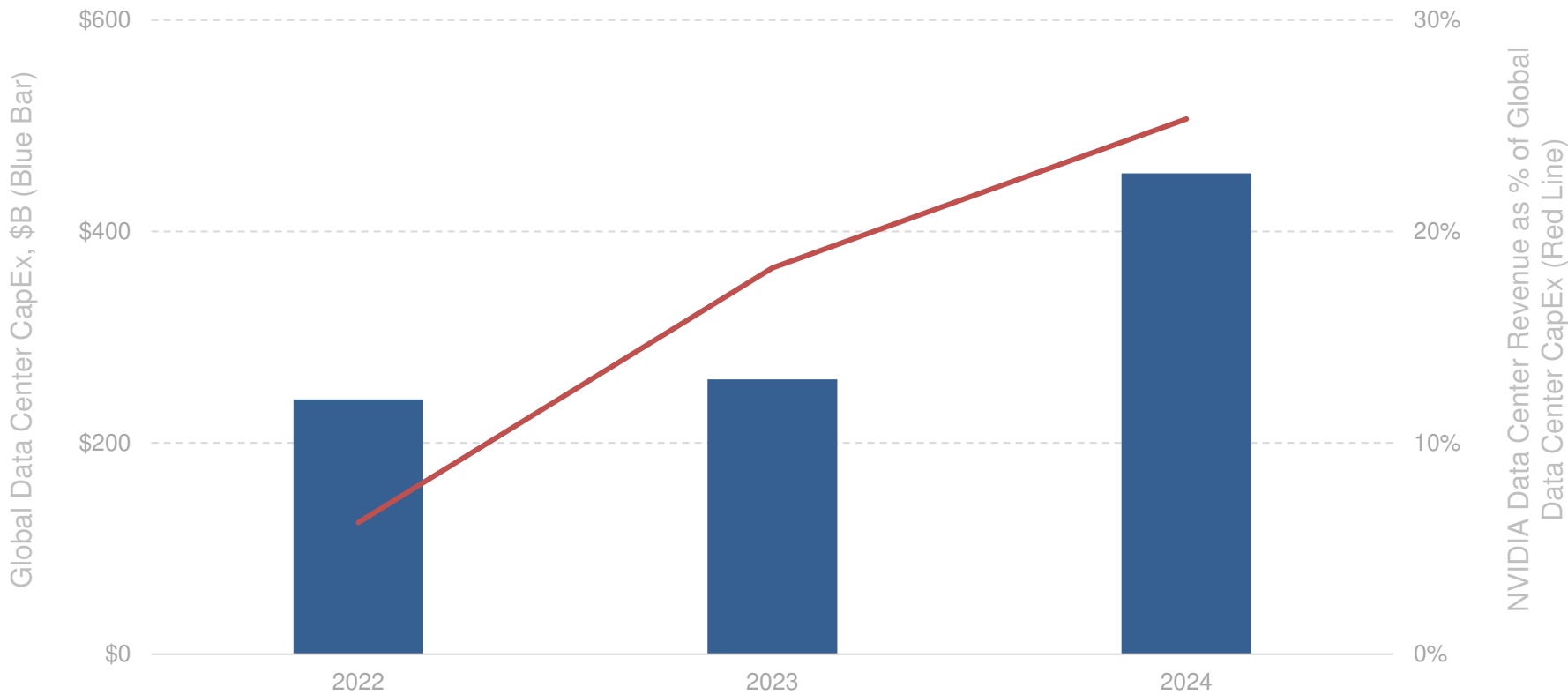
NVIDIA

Tech CapEx Spend Beneficiary =

NVIDIA

Key Tech CapEx Spend Beneficiary = NVIDIA...  
25% & Rising of Global Data Center CapEx, per NVIDIA

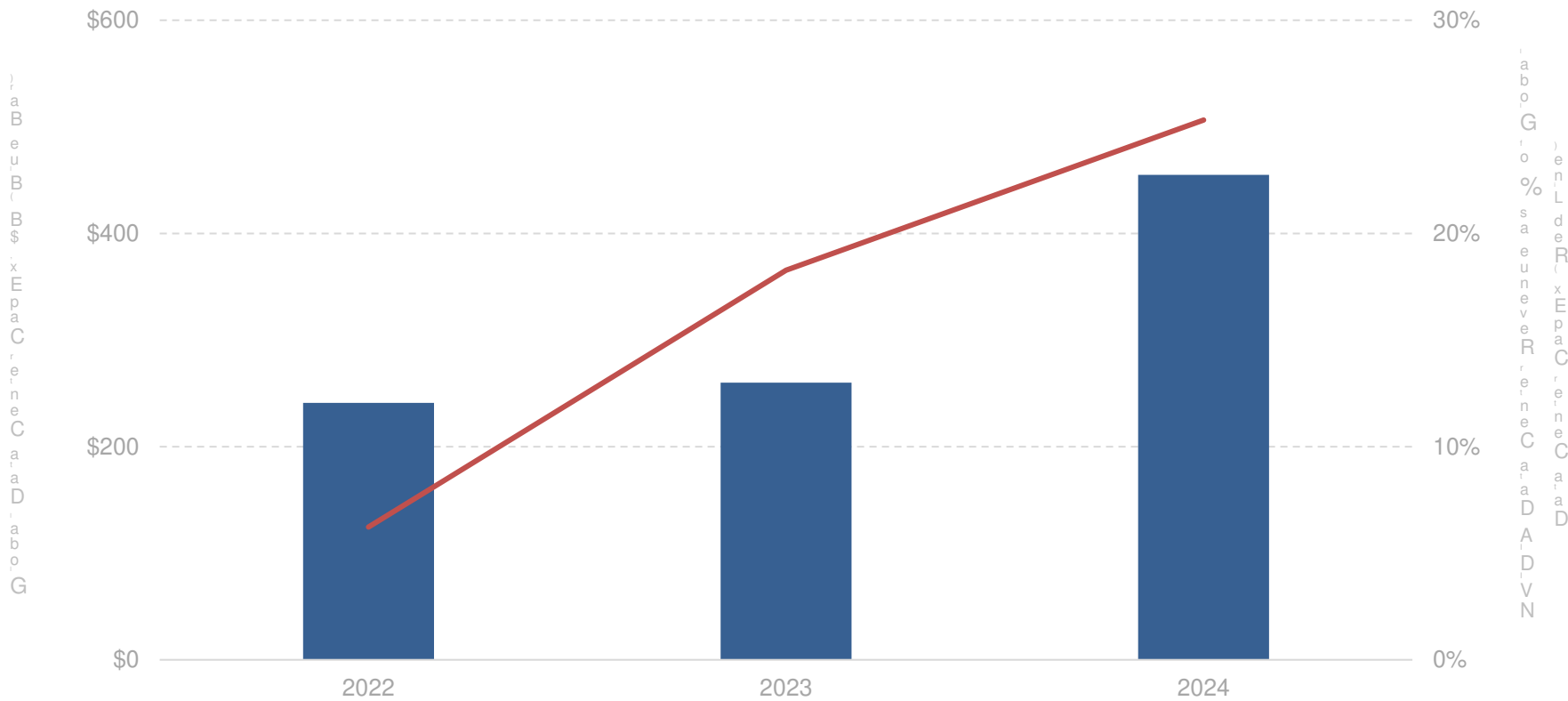
Global Data Center CapEx (\$B) vs. NVIDIA's Data Center Revenue as Percent of Data Center CapEx (Global) – 2022-2024, per NVIDIA @ GTC



Note: NVIDIA data represents January FYE (e.g., 2024 = FY25 ending 1/25) vs calendar year for data center CapEx. Data presented by Jensen Huang at NVIDIA GTC 2025 ([link](#)).  
Source: Dell'Oro Research for CapEx (3/25); NVIDIA for data center revenue (3/25)

主要技术资本支出受益者 = NVIDIA... 占全球数据中心资本支出  
出的 25% 且还在上升（根据 NVIDIA 数据）

全球数据中心资本支出（美元）与 NVIDIA 的数据中心收入占数据中心资本支出  
（全球）的百分比 – 2022-2024 年（根据 NVIDIA @ GTC 数据）



注：NVIDIA 数据代表 1 月份的财政年度末（例如，2024 = FY25 截止 1/25），而数据中心资本支出为日历年度。数据由 Jensen Huang 在 NVIDIA GTC 2025 上提供（链接）。来源：Dell'Oro Research for CapEx (3/25)；NVIDIA for data center revenue (3/25)

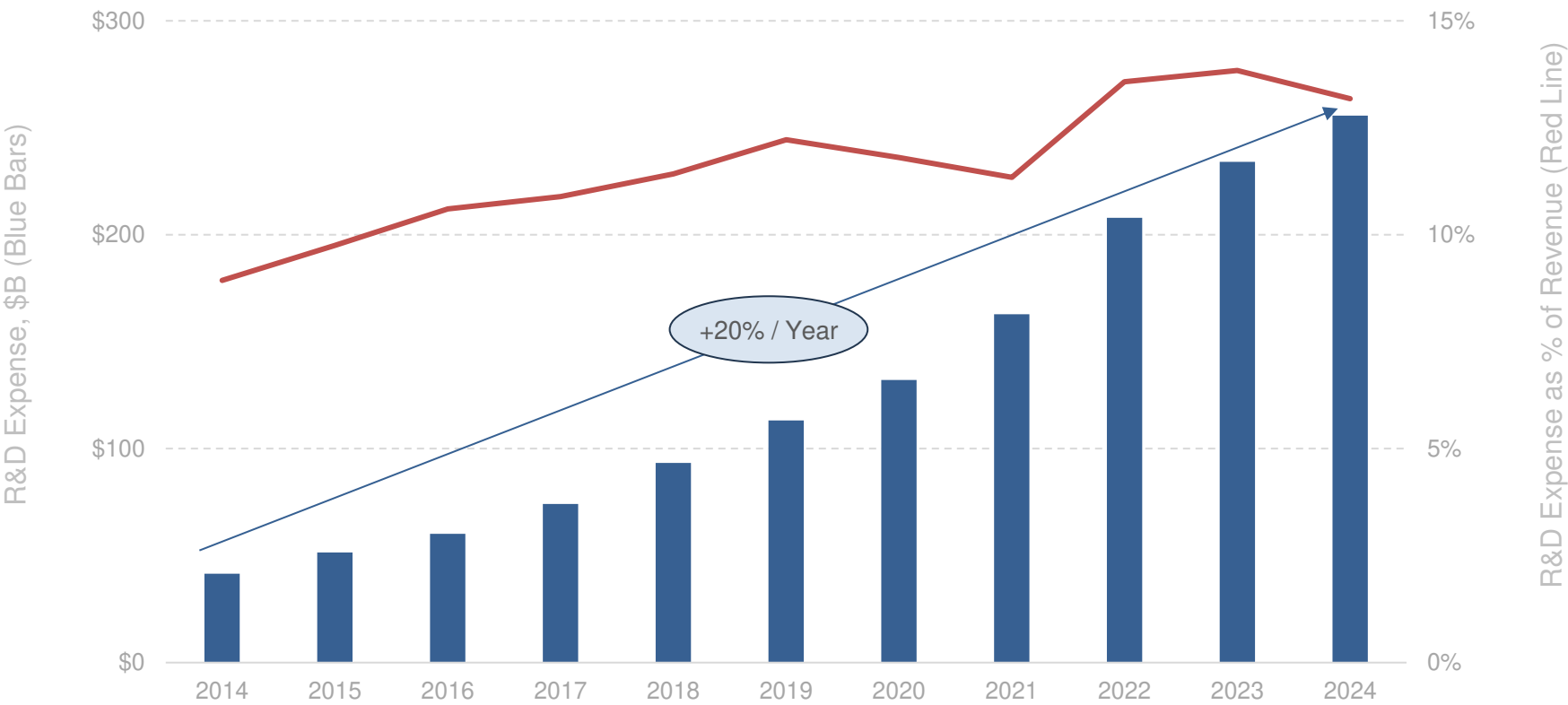
Technology Company Spend =  
R&D Rising Along with CapEx

科技公司支出 =  
研发支出随着资本支出增加

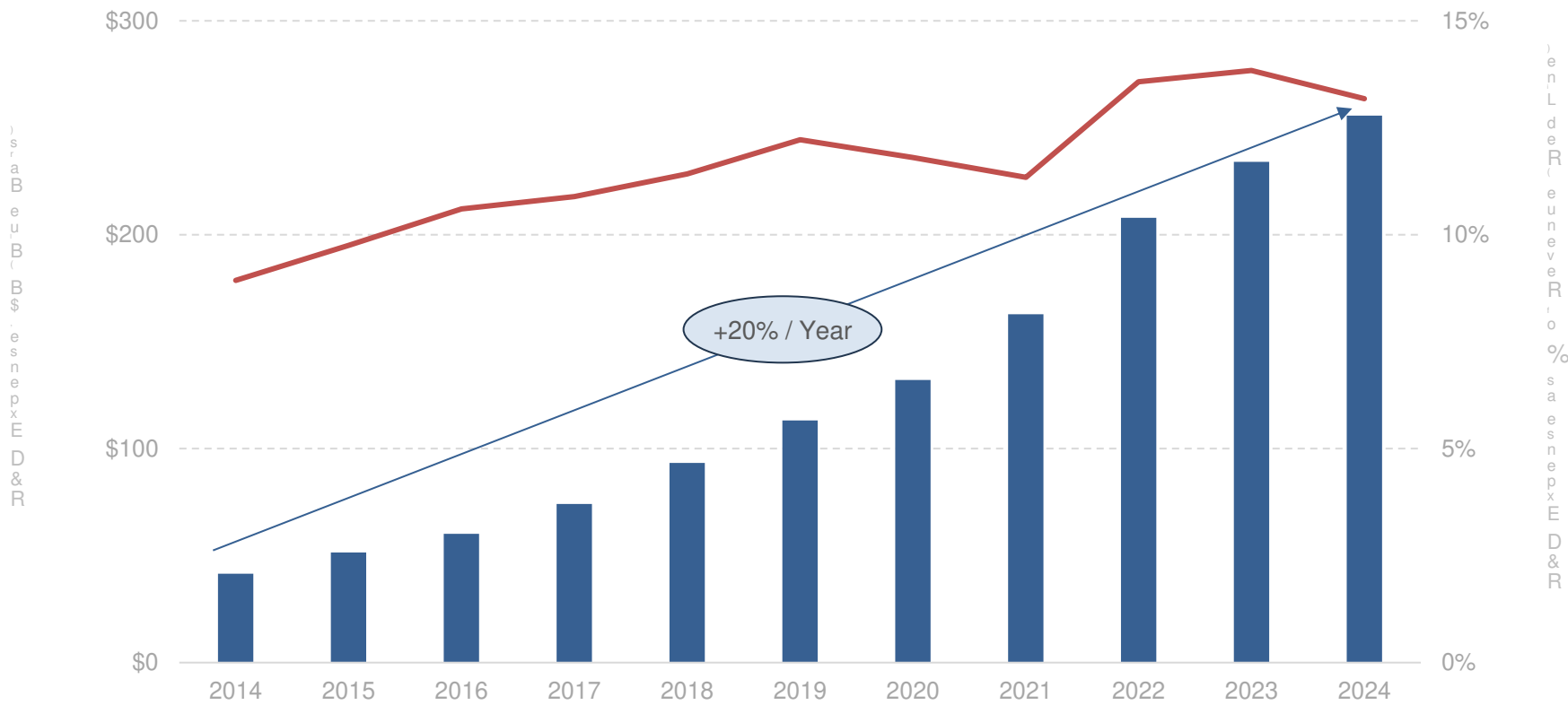
R&D Spend @ Big Six\* USA Public Tech Companies = 13% of Revenue...vs. 9% Ten Years Ago

美国六大 \* 上市科技公司研发支出 = 占收入的 13%，而十年前为 9%

Big Six\* USA Public Technology Company – R&D Spend (\$B) vs. % of Revenue – 2014-2024, per Capital IQ



Big Six\* USA Public Technology Company – R&D Spend (\$B) vs. % of Revenue – 2014-2024, per Capital IQ



\*Note: Big Six USA technology companies include Apple, Nvidia, Microsoft, Alphabet / Google, Amazon, & Meta Platforms / Facebook. R&D expense shown for Amazon, not AWS, as figures are not broken out in company financials; revenue therefore shown on like-for-like basis. Source: Capital IQ (3/25)

\* 注：美国六大科技公司包括苹果、英伟达、微软、Alphabet / 谷歌、亚马逊和 Meta Platforms / Facebook。亚马逊显示的研发费用，而非 AWS，因为该公司财务报表中没有细分这些数字；因此，收入也以类似的方式显示。来源：Capital IQ (3/25)

Tech Big Six (USA) =

Loaded With Cash to Spend on AI & CapEx

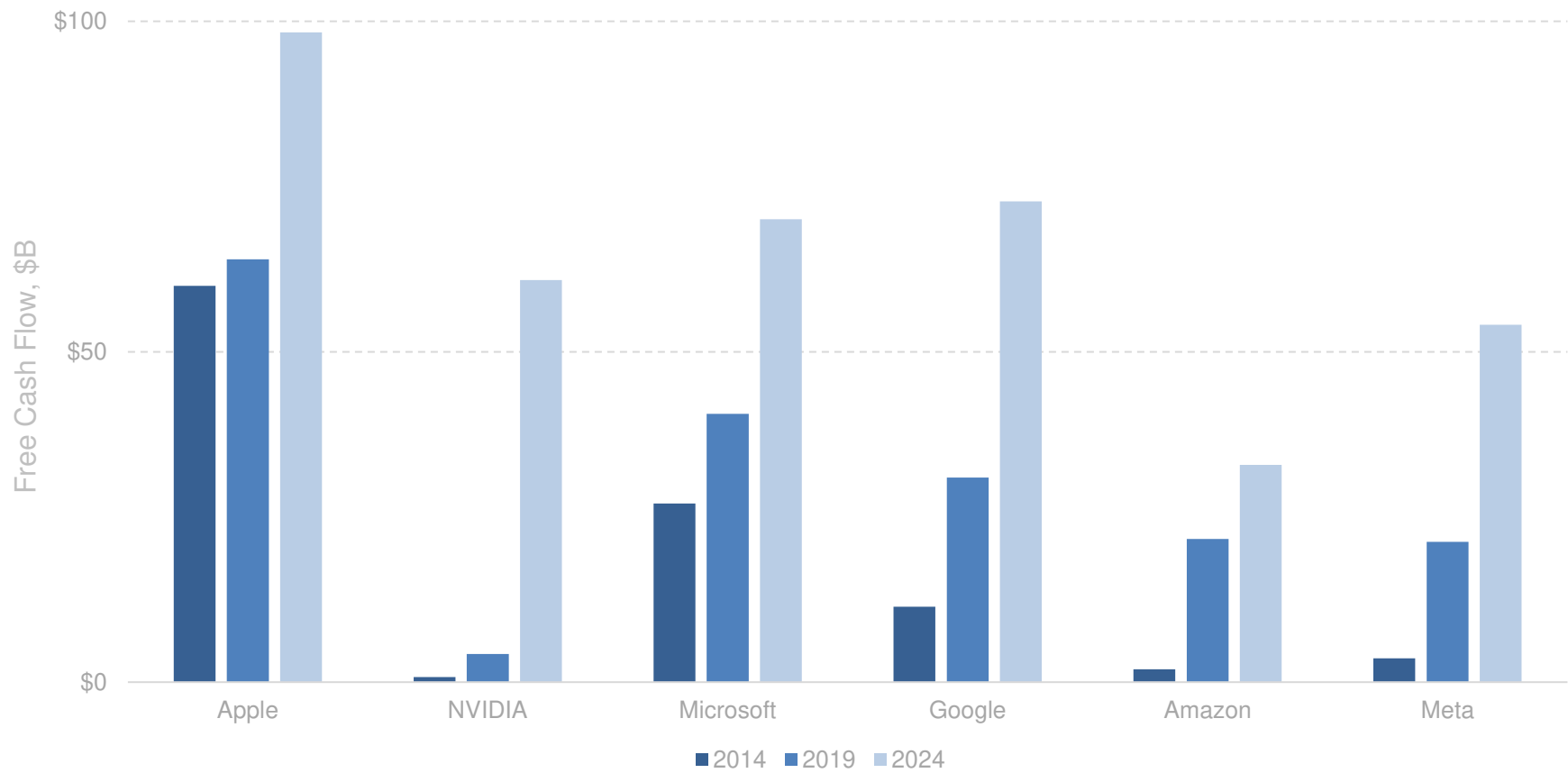
美国科技巨头六强 =

拥有大量现金用于人工智能和资本支出

Big Six\* Generating Loads of Cash =  
+263% Growth in Free Cash Flow Over Ten Years to \$389B...

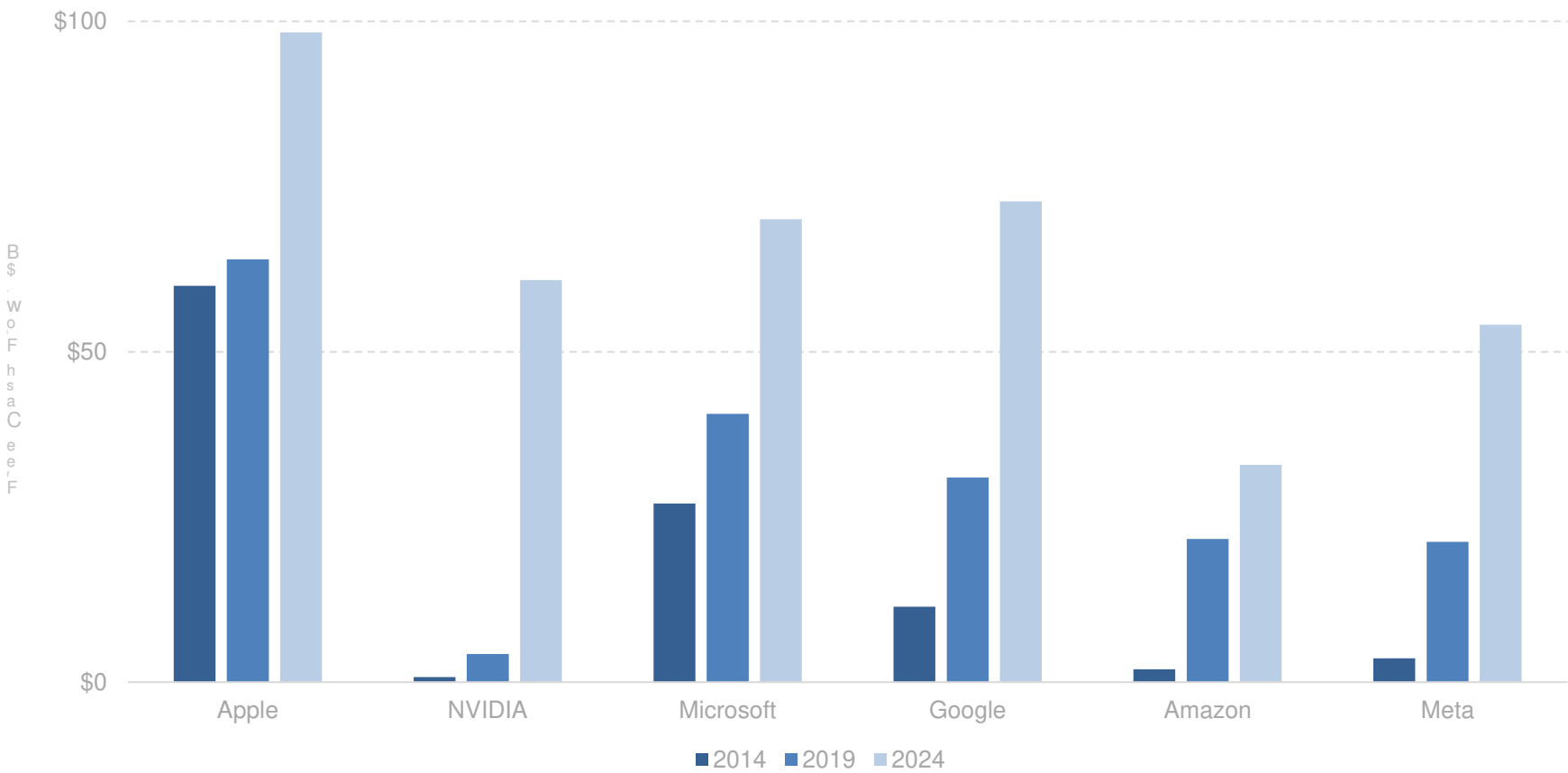
Big Six\* Generating Loads of Cash =+263% 过去十年自由现金流增长  
至 3890 亿美元 ……

Big Six\* Public Technology Companies – Free Cash Flow (\$B) – 2014-2024, per Capital IQ



*\*Note: Big Six USA technology companies include Apple, Nvidia, Microsoft, Alphabet / Google, Amazon, & Meta Platforms / Facebook. FCF calculated as cash flow from operations less capex to standardize definitions, as only some companies subtract finance leases and Amazon adjusts FCF for gains on sale of equipment. FCF shown for Amazon, not AWS, as figures are not broken out in company financials. Source: Capital IQ (3/25)*

Big Six\* Public Technology Companies – Free Cash Flow (\$B) – 2014-2024, per Capital IQ

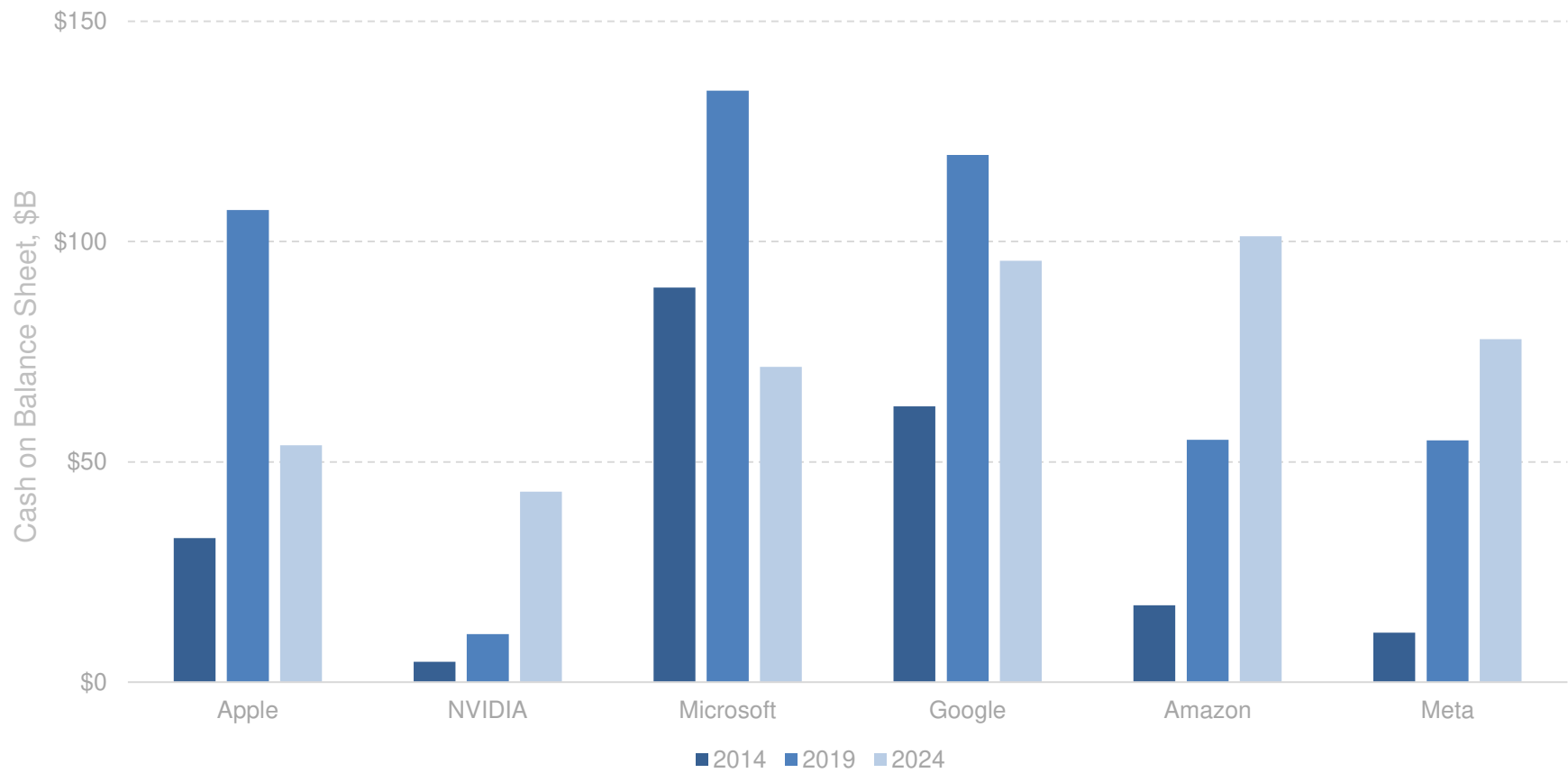


*\* 注：美国科技巨头六强包括苹果、英伟达、微软、Alphabet / 谷歌、亚马逊和 Meta Platforms / Facebook。自由现金流的计算方法为运营现金流减去资本支出，以标准化定义，因为只有部分公司扣除了融资租赁，而亚马逊调整了出售设备的收益的自由现金流。自由现金流显示的是亚马逊，而不是 AWS，因为这些数据没有在公司财务报表中单独列出。来源：Capital IQ (3/25)*

...Big Six\* Generating Loads of Cash =  
+103% Growth in Cash Over Ten Years to \$443B

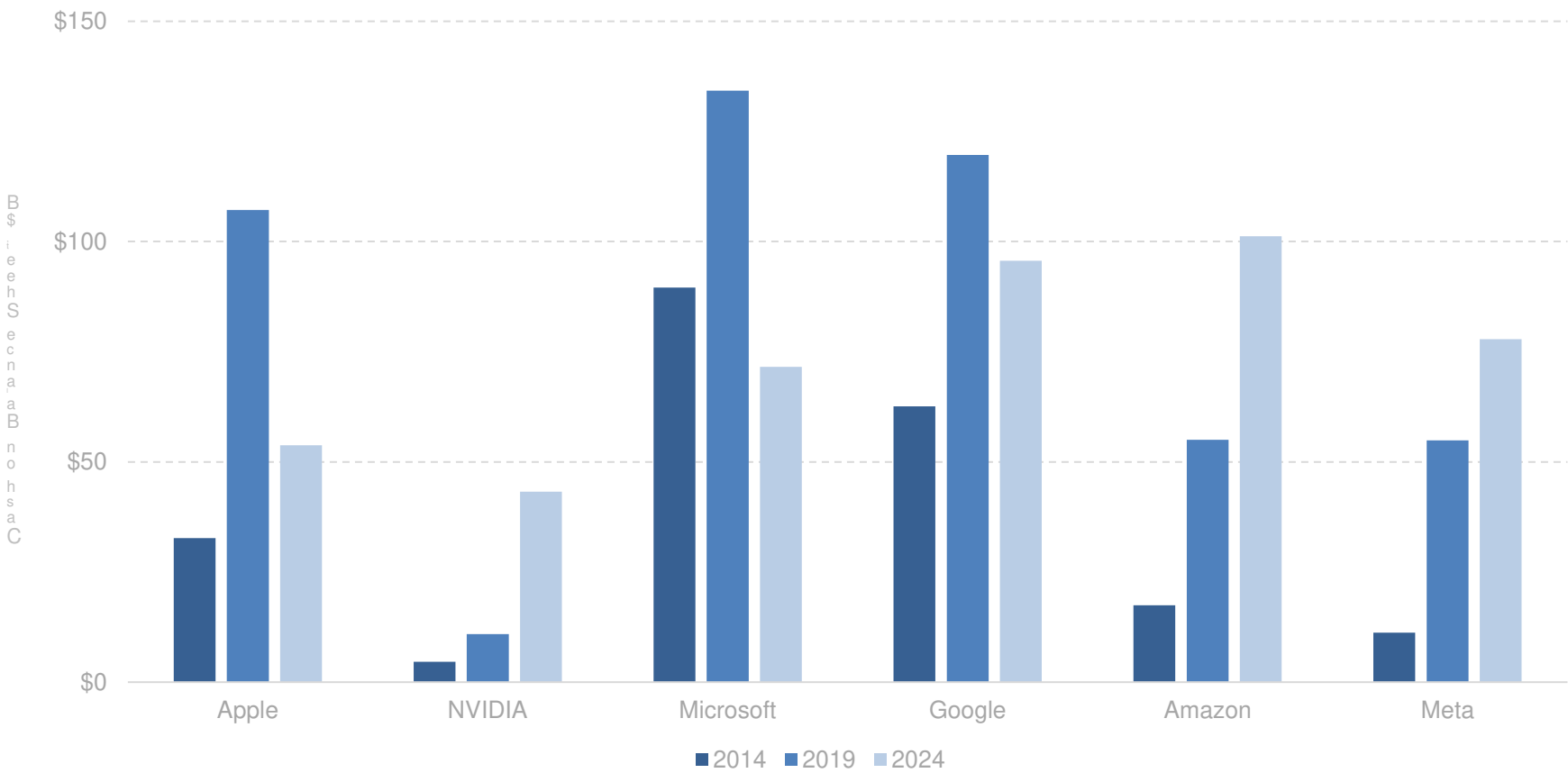
… 六巨头 \* 产生大量现金 =+103% 十年现金增长至  
4430 亿美元

Big Six\* USA Public Technology Company Cash on Balance Sheet (\$B) – 2014-2024, per Capital IQ



\*Note: Big Six USA technology companies include Apple, Nvidia, Microsoft, Alphabet / Google, Amazon, & Meta Platforms / Facebook. Figure measures cash and other equivalents (e.g., short-term investments and marketable securities) on companies' balance sheets. Source: Capital IQ (3/25)

美国六大 \* 上市科技公司资产负债表上的现金（十亿美元） – 2014-2024 年，数据来源：Capital IQ



\* 注：美国六大科技公司包括苹果、英伟达、微软、Alphabet / 谷歌、亚马逊和 Meta Platforms / Facebook。该数字衡量的是公司资产负债表上的现金和其他等价物（例如，短期投资和有色证券）。来源：Capital IQ (3/25)

Tech CapEx Spend Driver =

Compute Spend to Train & Run AI Models

技术资本支出驱动因素 =

用于训练和运行 AI 模型的计算支出



To understand the evolution of AI computing economics, it’s constructive to look at where costs are concentrated – And where they’re headed. The bulk of spending in AI large language model (LLM) development is still dominated by compute – specifically, the compute needed to train and run models. Training costs remain extraordinarily high and are rising fast, often exceeding \$100 million per model today. As Dario Amodei, CEO of Anthropic, noted in mid-2024, *Right now, [AI model training costs] \$100 million. There are models in training today that are more like a billion... I think that the training of...\$10 billion models, yeah, could start sometime in 2025.*

Around these core compute costs sit additional high-cost layers: research, data acquisition and hosting, and a mix of salaries, general overhead, and go-to-market operations. Even as the cost to train models climbs, a growing share of total AI spend is shifting toward inference – the cost of running models at scale in real-time. Inference happens constantly, across billions of prompts, queries, and decisions, whereas model training is episodic. As Amazon CEO Andy Jassy noted in his April 2025 letter to shareholders, *While model training still accounts for a large amount of the total AI spend, inference... will represent the overwhelming majority of future AI cost because customers train their models periodically but produce inferences constantly.* NVIDIA Co-Founder & CEO Jensen Huang noted the same in NVIDIA’s FQ1:26 earnings call, saying *Inference is exploding. Reasoning AI agents require orders of magnitude more compute.*

At scale, inference becomes a persistent cost center – one that grows in parallel with usage, despite declines in unit inference costs.

The broader dynamic is clear: lower per-unit costs are fueling higher overall spend. As inference becomes cheaper, AI gets used more. And as AI gets used more, total infrastructure and compute demand rises – dragging costs up again. The result is a flywheel of growth that puts pressure on cloud providers, chipmakers, and enterprise IT budgets alike.

The economics of AI are evolving quickly – but for now, they remain driven by heavy capital intensity, large-scale infrastructure, and a race to serve exponentially expanding usage.

To understand the evolution of AI computing economics, it’s constructive to look at where costs are concentrated – And where they’re headed. The bulk of spending in AI large language model (LLM) development is still dominated by compute – specifically, the compute needed to train and run models. Training costs remain extraordinarily high and are rising fast, often exceeding \$100 million per model today. As Dario Amodei, CEO of Anthropic, noted in mid-2024, *Right now, [AI model training costs] \$100 million. There are models in training today that are more like a billion... I think that the training of...\$10 billion models, yeah, could start sometime in 2025.*

Around these core compute costs sit additional high-cost layers: research, data acquisition and hosting, and a mix of salaries, general overhead, and go-to-market operations. Even as the cost to train models climbs, a growing share of total AI spend is shifting toward inference – the cost of running models at scale in real-time. Inference happens constantly, across billions of prompts, queries, and decisions, whereas model training is episodic. As Amazon CEO Andy Jassy noted in his April 2025 letter to shareholders, *While model training still accounts for a large amount of the total AI spend, inference... will represent the overwhelming majority of future AI cost because customers train their models periodically but produce inferences constantly.* NVIDIA Co-Founder & CEO Jensen Huang noted the same in NVIDIA’s FQ1:26 earnings call, saying *Inference is exploding. Reasoning AI agents require orders of magnitude more compute.*

At scale, inference becomes a persistent cost center – one that grows in parallel with usage, despite declines in unit inference costs.

更广泛的动态是明确的：单位成本降低正在推动更高的总体支出。随着推理成本降低，人工智能的使用量增加。随着人工智能的使用量增加，基础设施和计算总需求也在上升 – 再次推高成本。结果是一个增长的飞轮，给云提供商、芯片制造商和企业 IT 预算带来压力。

人工智能的经济学正在迅速发展 – 但目前，它们仍然受到重资本密集型、大规模基础设施以及服务于指数级增长的使用的竞赛的驱动。

Data Centers =

Key Beneficiary of AI CapEx Spend

数据中心 =

AI 资本支出重点受益者

For one lens into the economics of AI infrastructure,  
it's useful to look at the pace and scale of data center construction.  
The current wave of AI-driven demand has pushed data center spending to historic highs.  
According to Dell'Oro Research, global IT company data center CapEx  
reached \$455 billion in 2024 and is accelerating.

Hyperscalers and AI-first companies alike are pouring billions into building out  
compute-ready capacity – not just for storage, but for real-time inference and  
model training workloads that require dense, high-power hardware.  
As AI moves from experimental to essential, so too do data centers.  
Per NVIDIA Co-Founder and CEO Jensen Huang, *These AI data centers...are, in fact, AI factories.*

That race is moving faster than many expected.  
The most striking example may be xAI's Colossus facility in Memphis, Tennessee which went  
from a gutted factory to a fully operational AI data center in just 122 days.  
As noted on page 122, at 750,000 square feet – roughly the size of 418 average USA homes –  
it was built in half the time it typically takes to construct a single American house.

Per NVIDIA Co-Founder & CEO Jensen Huang,  
*What they achieved is singular, never been done before...That is, like, superhuman...*

要了解人工智能基础设施的经济状况，查看数据中心建设的速度和规模很有用。当前  
人工智能驱动的需求浪潮已将数据中心支出推至历史新高。根据 Dell'Oro Research 的数  
据，2024 年全球 IT 公司数据中心的资本支出达到 4550 亿美元，并且还在加速增长。

超大规模企业和以人工智能为先的公司都在投入数十亿美元来构建可用于计算的容量 —— 不仅  
用于存储，还用于需要密集型、高功率硬件的实时推理和模型训练工作负载。随着人工智能从实验  
性走向必要性，数据中心也随之发展。正如 NVIDIA 联合创始人兼首席执行官黄仁勋所说，这些人  
工智能数据中心 …… 实际上就是人工智能工厂。

这场竞赛的进展速度超出了许多人的预期。最引人注目的例子可能是 xAI 的位于田纳西州  
孟菲斯的 Colossus 工厂，该工厂仅用了 122 天就从一个破败的工厂变成了一个完全运营的人  
工智能数据中心。正如第 122 页所述，该工厂占地 75 万平方英尺 —— 大约相当于 418 个美国  
普通住宅的大小 —— 它的建造时间只有建造一栋美国普通房屋所需时间的一半。

根据 NVIDIA 联合创始人兼首席执行官黄仁勋的说法，他们取得的成就是独一无二的，前  
所未有 …… 就像，超人一样 ……

...These kinds of timelines are no longer the exception. With prefabricated modules, streamlined permitting, and vertical integration across electrical, mechanical, and software systems, new data centers are going up at speeds that resemble consumer tech cycles more than real estate development.

But beneath that velocity lies a capital model that’s anything but simple. CapEx is driven by land, power provisioning, chips, and cooling infrastructure – especially as AI workloads push thermal and power limits far beyond traditional enterprise compute. OpEx, by contrast, is dominated by energy costs and systems maintenance, particularly for high-density training clusters that operate near constant load.

Revenue is driven by compute sales – whether in the form of AI APIs, enterprise platform fees, or internal productivity gains. But payback periods are often long, especially for vertically-integrated players building ahead of demand. For newer entrants, monetization may lag build-out by quarters or even years.

And then there’s the supply chain. Power availability is becoming more of a gating factor. Transformers, substations, turbines, GPUs, cables – these aren’t commodities that can be spun up overnight. In this context, data centers aren’t just physical assets – they are strategic infrastructure nodes. They sit at the intersection of real estate, power, logistics, compute, and software monetization.

The companies that get this right may do more than run servers – they will shape the geography of AI economics for the next decade.

… 这些时间表不再是例外。借助预制模块、简化的许可流程以及电气、机械和软件系统的垂直整合，新建数据中心的速度比房地产开发更像是消费科技周期。

但在这种速度的背后，隐藏着一种绝不简单的资本模型。资本支出由土地、电力供应、芯片和冷却基础设施驱动 – 特别是当人工智能工作负载将热力和功率限制远远超出传统企业计算时。相比之下，运营支出主要由能源成本和系统维护构成，特别是对于以接近恒定负载运行的高密度训练集群。

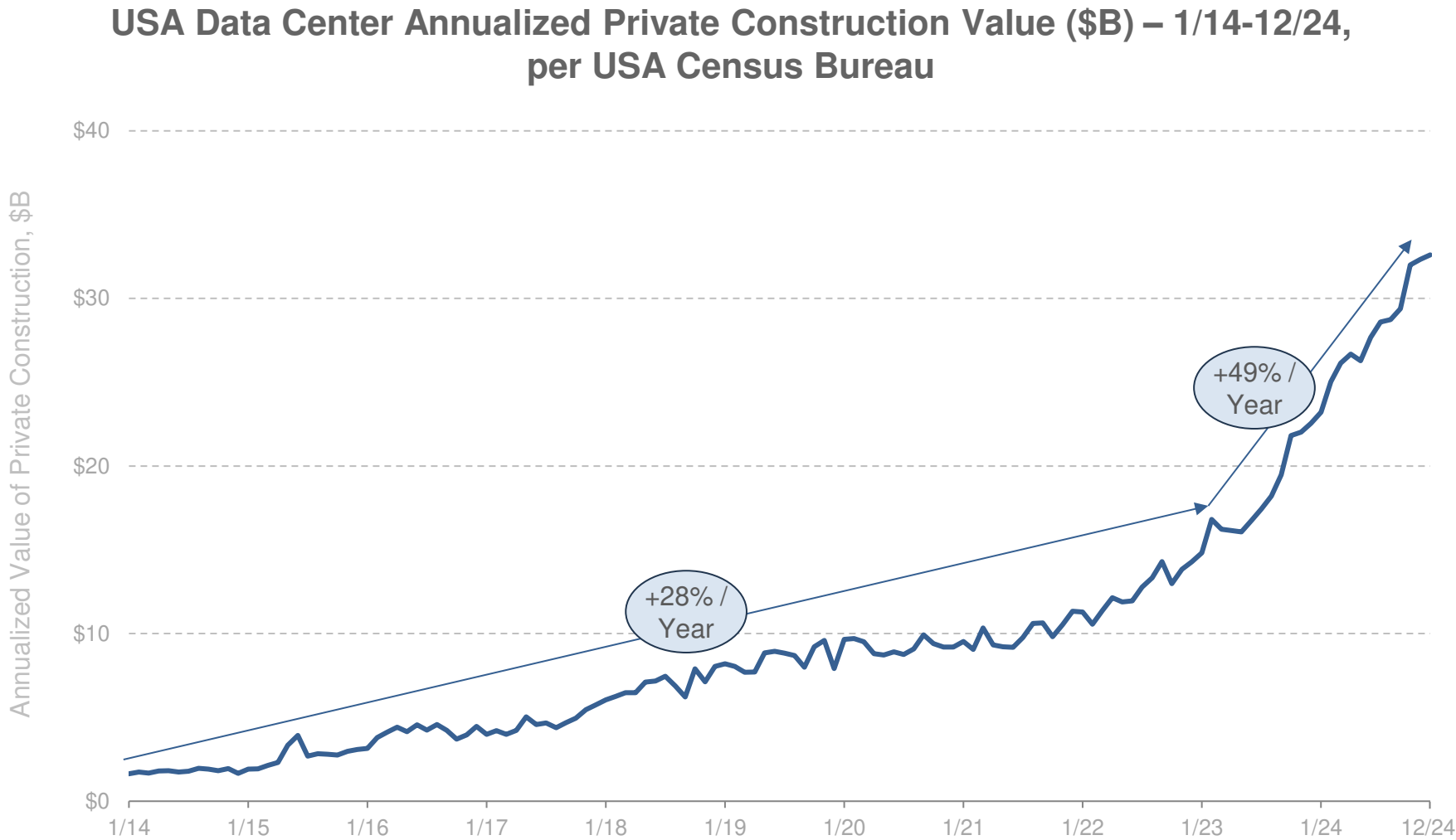
收入由计算销售驱动 – 无论是人工智能 API、企业平台费用还是内部生产力提升。但投资回收期通常很长，尤其是对于在需求之前进行垂直整合的参与者。对于新进入者来说，货币化可能会比建设滞后几个季度甚至几年。

还有供应链的问题。电力供应正日益成为一个制约因素。变压器、变电站、涡轮机、GPU、电缆 —— 这些不是可以一夜之间启动的商品。在这种背景下，数据中心不仅仅是物理资产 —— 它们是战略基础设施节点。它们位于房地产、电力、物流、计算和软件货币化的交汇点。

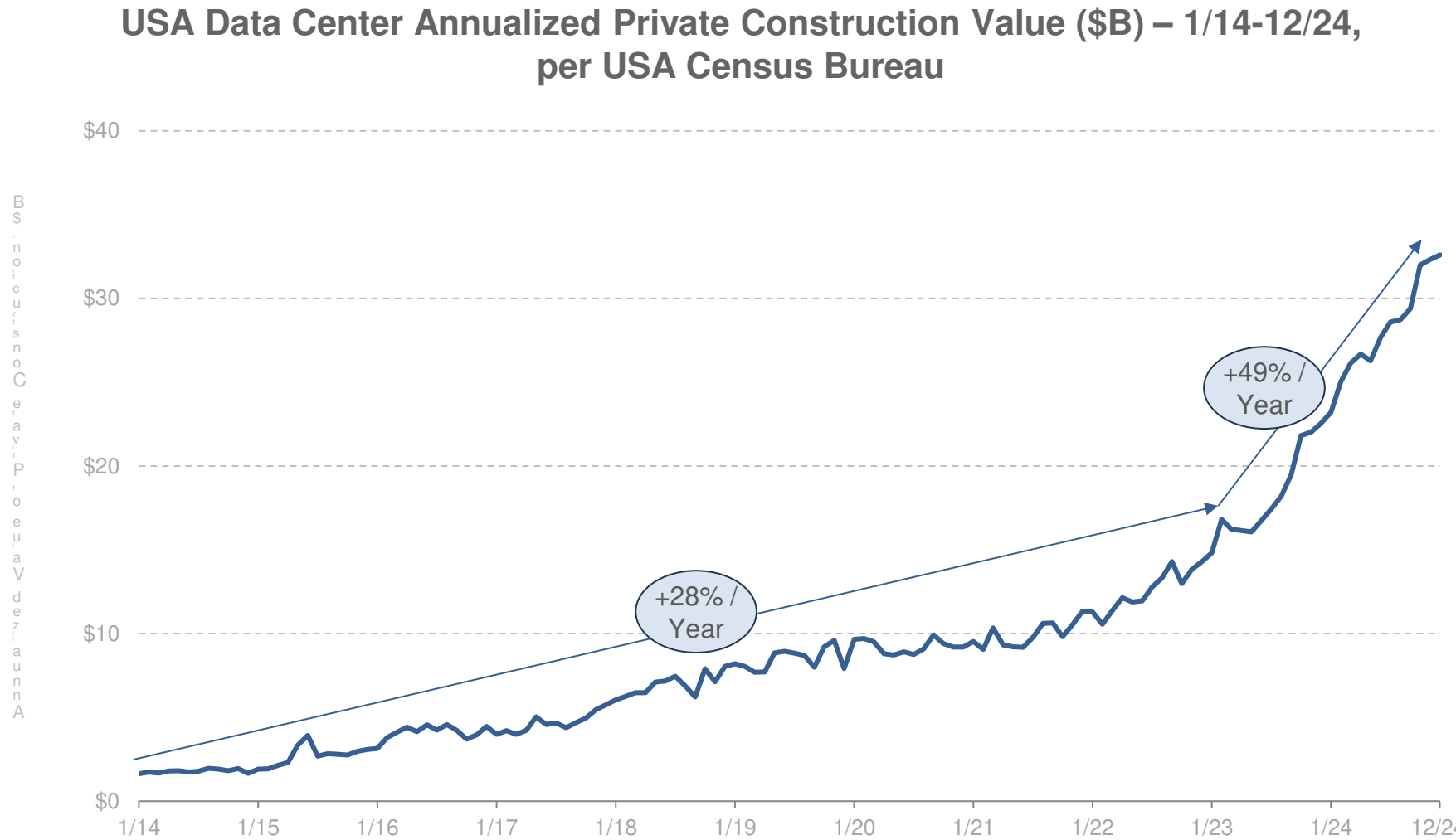
能够正确理解这一点的公司，可能不仅仅是运行服务器 —— 它们将在未来十年内塑造人工智能经济的地理格局。

Data Center Buildout Construction Value, USA = +49% & Accelerated Annual Growth Over Two Years

Data Center Buildout Construction Value, USA = +49% & Accelerated Annual Growth Over Two Years



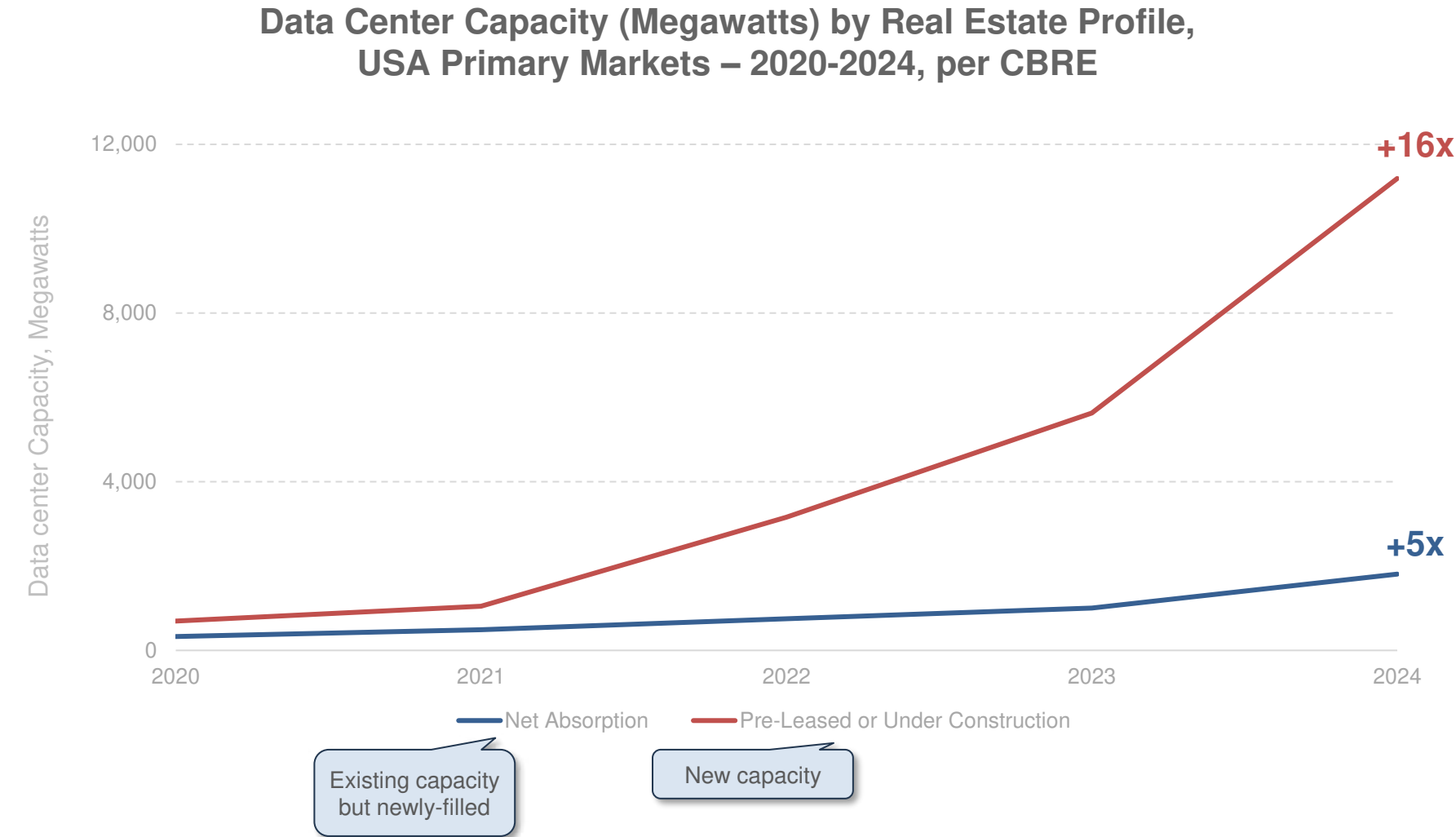
Note: All data are seasonally adjusted. Data obtained via USA Census Bureau's Value of Construction Put in Place (VIP) Survey, which provides monthly estimates of the total dollar value of construction work done in USA. Data is annualized to avoid seasonal fluctuations. Source: USA Census Bureau



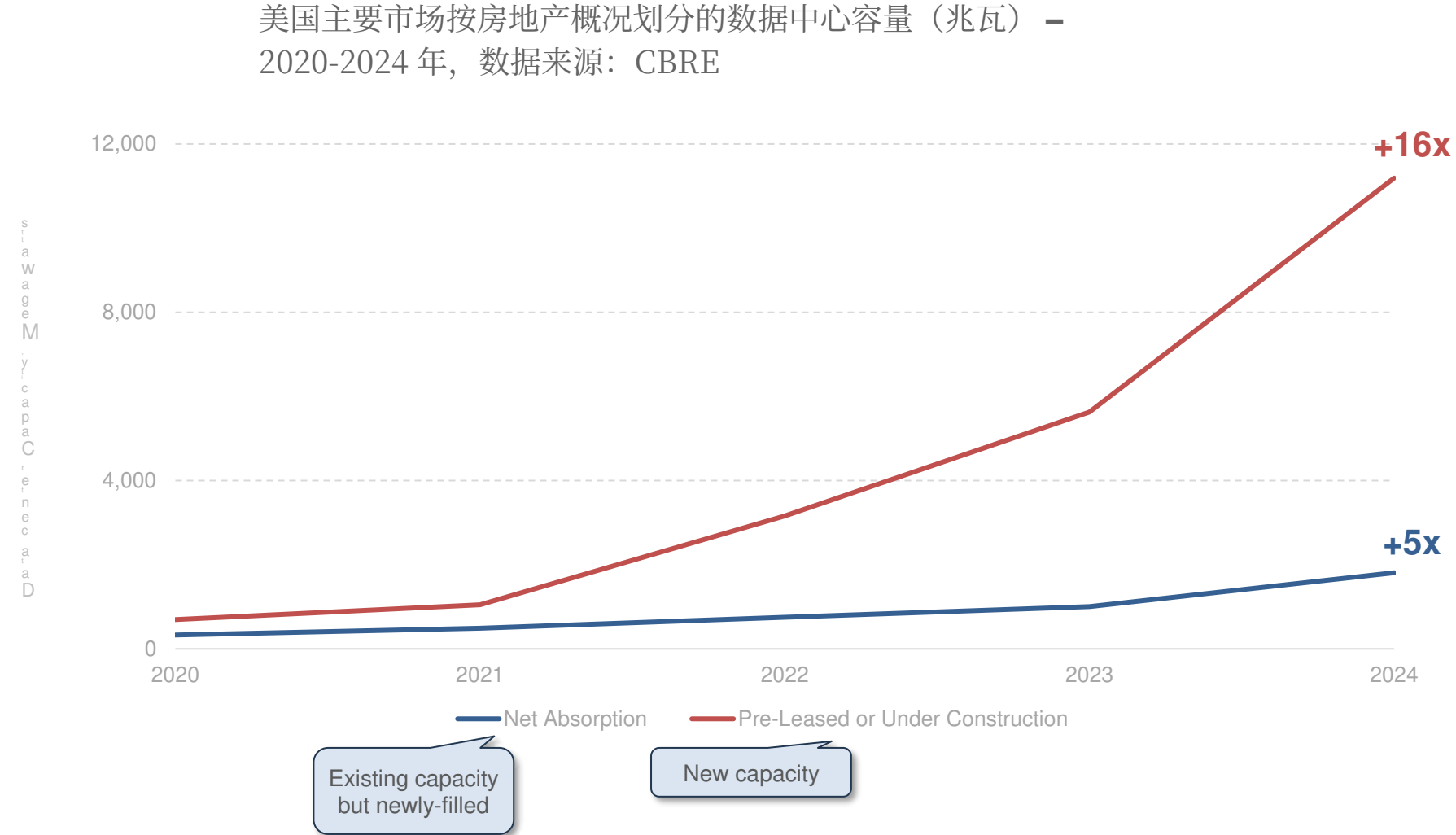
注意：所有数据均经过季节性调整。数据来自美国人口普查局的在建工程价值 (VIP) 调查，该调查提供了美国完成的建筑工程总价值的月度估算。数据已进行年化处理，以避免季节性波动。来源：美国人口普查局

Data Center New Construction vs. Existing Capacity, USA = +16x in New vs. +5x in Existing Over Four Years

美国数据中心新建与现有容量对比 =+16 新建容量是现有容量的 +5 倍（四年期间）



Note: Primary USA markets only (Northern Virginia, Atlanta, Chicago, Phoenix, Dallas-Ft. Worth, Hillsboro, Silicon Valley, New York Tri-State.)   
Source: CBRE, "North America Data Center Trends H2 2024" (2/25)



注：仅限美国主要市场（北弗吉尼亚、亚特兰大、芝加哥、凤凰城、达拉斯-沃斯堡、希尔斯伯勒、硅谷、纽约三州）。来源：CBRE, "2024 年下半年北美数据中心趋势" (2/25)



Data Center Build Time (xAI Colossus as Proxy) =  
122 Days vs. 234 for a Home

数据中心建设时间（ xAI Colossus 作为代理） =122 天  
vs. 住宅 234 天

122 Days =  
A Fully-Operational Data Center – 2024...  
750,000 Sq. Ft = Size of 418 USA Homes



750,000 Square Feet

We were told it would take 24 months to build. So we took the project into our own hands, questioned everything, removed whatever was unnecessary, and accomplished our goal in four months.

- xAI Website

122 Days =  
One Half-Built House – 2024  
(Average Build Time = 234 Days)



1,792 Square Feet

122 天 = 一个完全运行的数据中心 – 2024...  
750,000 平方英尺 = 相当于 418 个美国住宅的面积



750,000 平方英尺

我们被告知需要 24 个月才能建成。因此，我们亲自接手了这个项目，质疑一切，去除一切不必要的，并在四个月内实现了我们的目标。

- xAI 网站

122 Days =  
一栋半建成的房子 – 2024  
(平均建设时间 = 234 天)



1,792 平方英尺

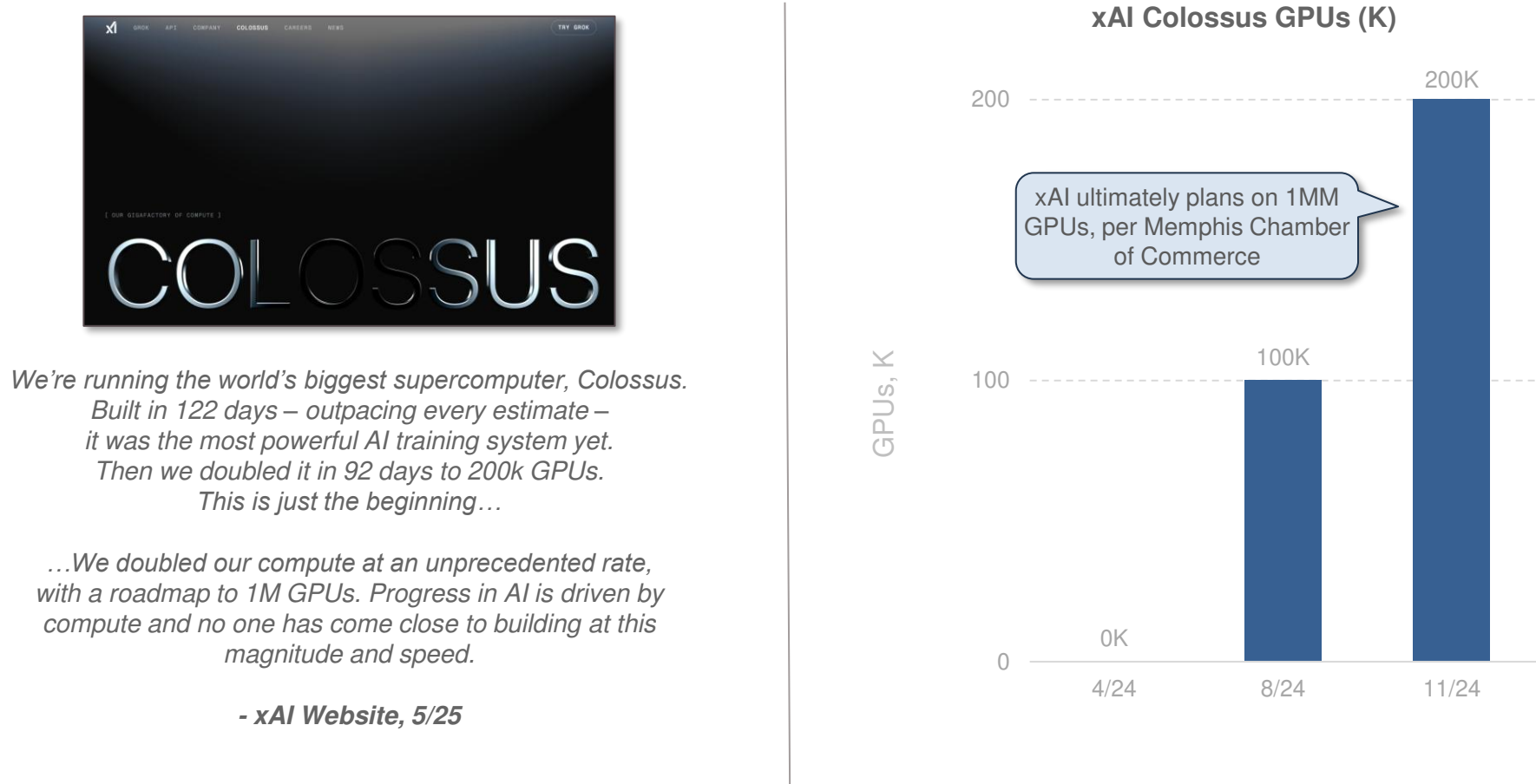
Note: Median USA home size shown as of January 2025, per FRED. Colossus was built in a former Electrolux factory in Memphis, TN, USA. Average build time shown for single-unit buildings. Measures time between start of onsite work & completion. Data reported in 2024 but measures build times for homes started in 2023. Source: xAI, USA Census Bureau, Federal Reserve Bank of St. Louis, Wikimedia Commons

注意：根据 FRED，截至 2025 年 1 月，美国房屋中位数面积。Colossus 建于美国田纳西州孟菲斯市的前伊莱克斯工厂。显示单体建筑的平均建设时间。衡量现场工作开始到完成之间的时间。数据于 2024 年报告，但衡量的是 2023 年开始建造的房屋的建设时间。来源：xAI、美国人口普查局、圣路易斯联邦储备银行、维基共享资源

Data Center Compute (xAI Colossus as Proxy) = 0 to 200,000 GPUs in Seven Months

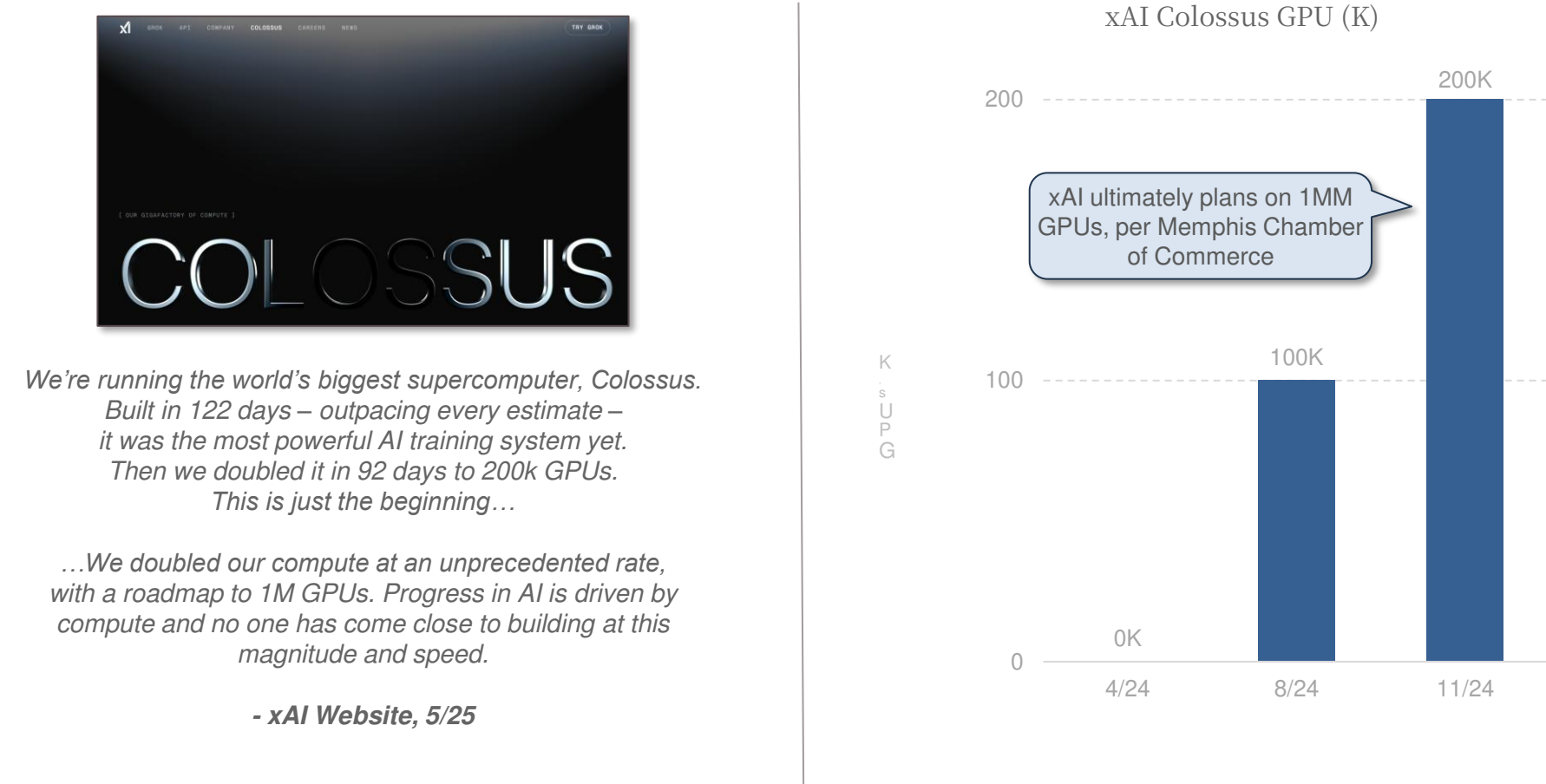
数据中心计算（ xAI Colossus 作为代理） = 在七个月内从 0 到 200,000 个 GPU

xAI Colossus GPUs – 4/24-11/24, per xAI



Note: We assume 200,000 GPUs as of 11/30/24 per xAI's disclosure that 'we doubled [GPU count] in 92 days to 200K GPUs.' xAI Colossus ran its first job across 4 data halls on 8/30/24. We assume zero GPUs as of construction start date (122 days prior to assumed opening date of 8/30/24). Source: xAI (5/25), Memphis Chamber of Commerce (12/24)

xAI Colossus GPU –4/24-11/24, 根据 xAI



注意：根据 xAI 披露的“我们在 92 天内将 [GPU 数量] 增加了一倍，达到 20 万个 GPU”，我们假设截至 11/30/24 有 20 万个 GPU。xAI Colossus 于 8/30/24 在 4 个数据大厅中运行了第一个作业。我们假设从施工开始日期（假定的 8/30/24 开工日期之前的 122 天）开始 GPU 数量为零。来源：xAI (5/25)，孟菲斯商会 (12/24)



Data Centers =  
Electricity Guzzlers

数据中心 =  
耗电大户

AI and energy observations / quotes (in italics) here and the two pages that follow are from ‘World Energy Outlook Special Report – Energy and AI’ (link) from IEA (International Energy Agency)\* – 4/10/25

To understand where energy infrastructure is heading, it helps to examine the rising tension between AI capability and electrical supply. The growing scale and sophistication of artificial intelligence is demanding an extraordinary amount of computational horsepower, primarily from AI-focused data centers. These facilities – purpose-built to train and serve models – are starting to rival traditional heavy industry in their electricity consumption.

*There is no AI without energy – specifically electricity (p. 3). Data centers accounted for around 1.5% of the world’s electricity consumption in 2024 (p. 14). Energy demand growth has been rapid: Globally, data centre electricity consumption has grown by around 12% per year since 2017, more than four times faster than the rate of total electricity consumption (p. 14). As power demand rises, so too does its concentration: The United States accounted for...[45% of global data centre electricity consumption], followed by China (25%) and Europe (15%)... nearly half of data centre capacity in the United States is in five regional clusters (p. 14).*

The flipside is true as well: *Emerging and developing economies other than China account for 50% of the world’s internet users but less than 10% of global data centre capacity (p. 18)...*

此处以及后续两页的 AI 和能源观察 / 引言（斜体字）均来自 “世界能源展望特别报告 —— 能源与人工智能” (link)，由 IEA （国际能源署）发布 \* – 4/10/25

要了解能源基础设施的发展方向，需要研究人工智能之间日益紧张的关系能力和电力供应。人工智能的日益增长的规模和复杂性需要大量的计算能力，主要来自以人工智能为中心的数据中心。这些设施 – 专门用于训练和服务模型 –，它们在电力消耗方面开始与传统的重工业相媲美。

没有能源就没有人工智能 – 特别是电力（第 3 页）。数据中心约占世界 2024 年电力消耗量的 1.5%（第 14 页）。能源需求增长迅速：自 2017 年以来，全球数据中心电力消耗量每年增长约 12%，是电力总消耗量增速的四倍多（第 14 页）。随着电力需求的增长，其集中度也在上升：美国占 ...[45% 的全球数据中心电力消耗量 ]，其次是中国（25%）和欧洲（15%） ... 美国近一半的数据中心容量位于五个区域集群中（第 14 页）。

另一方面也是如此：除中国以外的新兴和发展中经济体占世界互联网用户总数的 50%，但仅占全球数据中心容量的不到 10%（第 18 页） ...

...AI’s power demands are increasing – and its progress is increasingly bottlenecked not by data or algorithms, but by the grid and strains related to demand.

While AI presently places considerable demands on the energy sector, it is also already unlocking major energy efficiency and operational gains... *AI is already being deployed by energy companies to transform and optimize energy and mineral supply, electricity generation and transmission, and energy consumption (p. 16).*

Current AI-driven demand is extremely high. This is forecast to continue, especially as capital gushes into model providers that, in turn, spend on more compute. At some point, these model builders will need to turn a profit to be able to spend more.

While demand – for both compute and energy – will inevitably continue to rise as consumer and business usage does the same, data centers will ultimately only serve those who pay their bills.

\*IEA member countries include Australia, Austria, Belgium, Canada, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Japan, S. Korea, Latvia, Lithuania, Luxembourg, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Spain, Sweden, Switzerland, Republic of Türkiye, United Kingdom, and United States. IEA Association countries include Argentina, Brazil, China, Egypt, India, Indonesia, Kenya, Morocco, Senegal, Singapore, S. Africa, Thailand, and Ukraine.

*All data shown, unless otherwise specified, is global. Italicized text is directly quoted from the report.*

… 人工智能的电力需求正在增加 – 其进步越来越受到电网和与需求相关的压力的瓶颈，而不是数据或算法。

虽然人工智能目前对能源部门提出了相当大的需求，但它也已经在释放主要的能源效率和运营收益 …… 能源公司已经开始部署人工智能，以转变和优化能源和矿产供应、发电和输电以及能源消耗（第 16 页）。

目前人工智能驱动的需求非常高。预计这种情况将持续下去，尤其是在资金涌入模型提供商，而模型提供商又会将资金用于更多计算的情况下。在某个时候，这些模型构建者需要盈利才能能够花费更多。

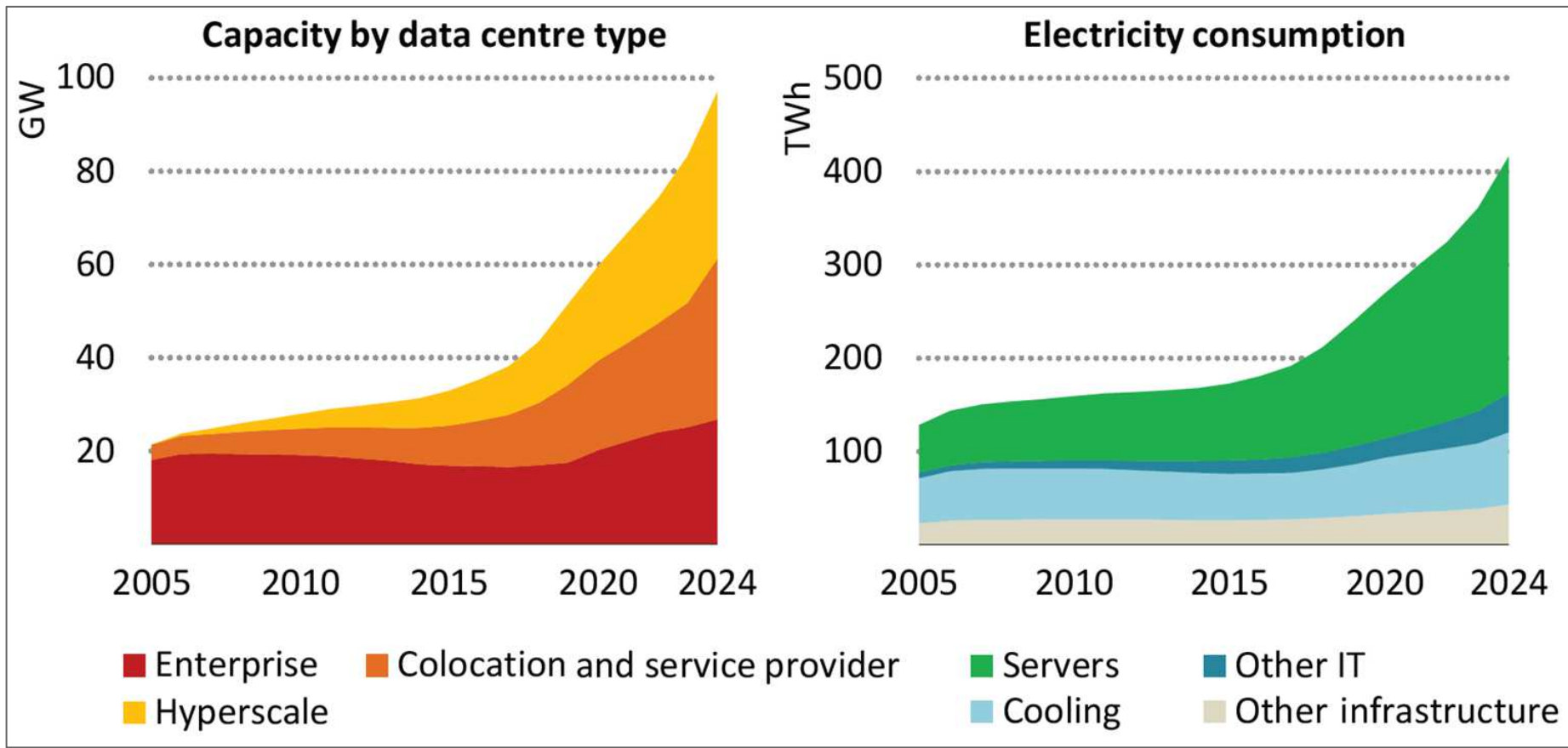
虽然对计算和能源的需求 – 将不可避免地随着消费者和企业的使用而继续上升 – ，但数据中心最终只会为那些支付账单的人提供服务。

\*IEA member countries include Australia, Austria, Belgium, Canada, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Japan, S. Korea, Latvia, Lithuania, Luxembourg, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Spain, Sweden, Switzerland, Republic of Türkiye, United Kingdom, and United States. IEA Association countries include Argentina, Brazil, China, Egypt, India, Indonesia, Kenya, Morocco, Senegal, Singapore, S. Africa, Thailand, and Ukraine.

除非另有说明，否则显示的所有数据均为全球数据。<i> 斜体文本直接引自报告。</i>

# Data Center Electricity Consumption, Global = +3x Over Nineteen Years, per IEA

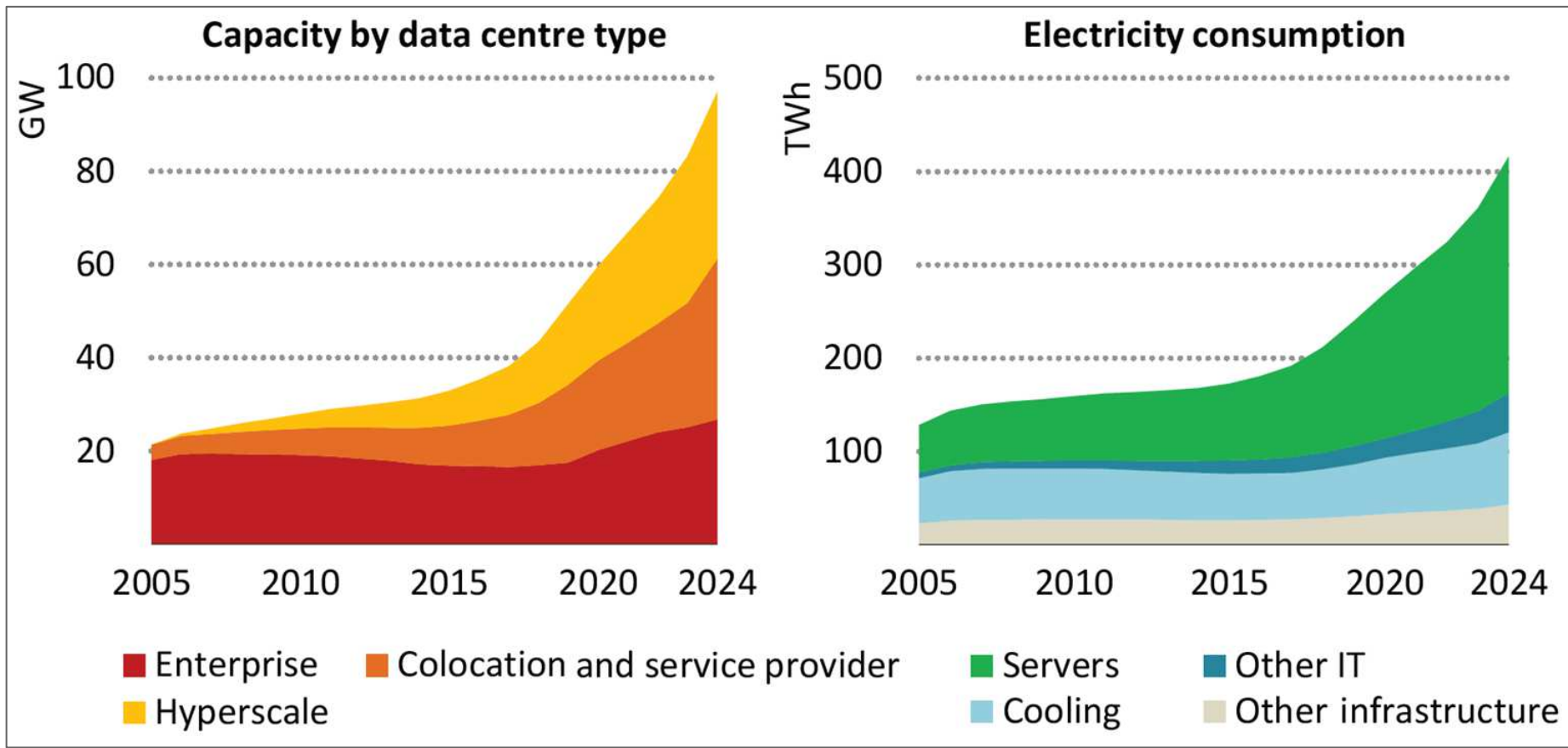
Data Center Energy Consumption by Data Center Type & Equipment, Global – 2005-2024, per IEA



Source: International Energy Agency (IEA), 'Energy and AI' (4/25)

# 全球数据中心用电量，根据 IEA，=+3 十九年内增长 x 倍

全球数据中心类型和设备的数据中心能耗 – 2005-2024 年（根据 IEA ）

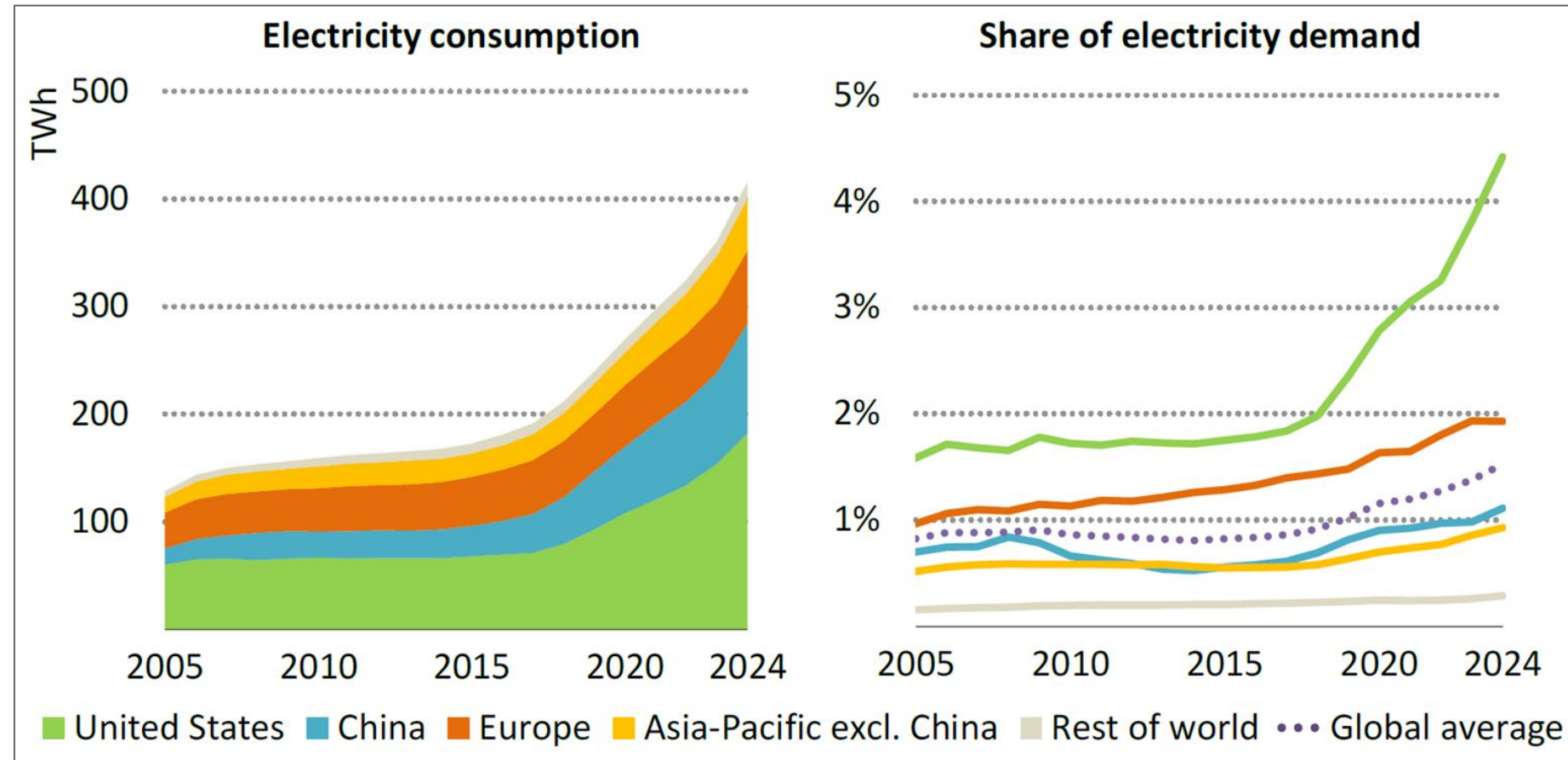


Source: International Energy Agency (IEA), 'Energy and AI' (4/25)

Data Center Electricity Consumption by Region = USA Leads, per IEA

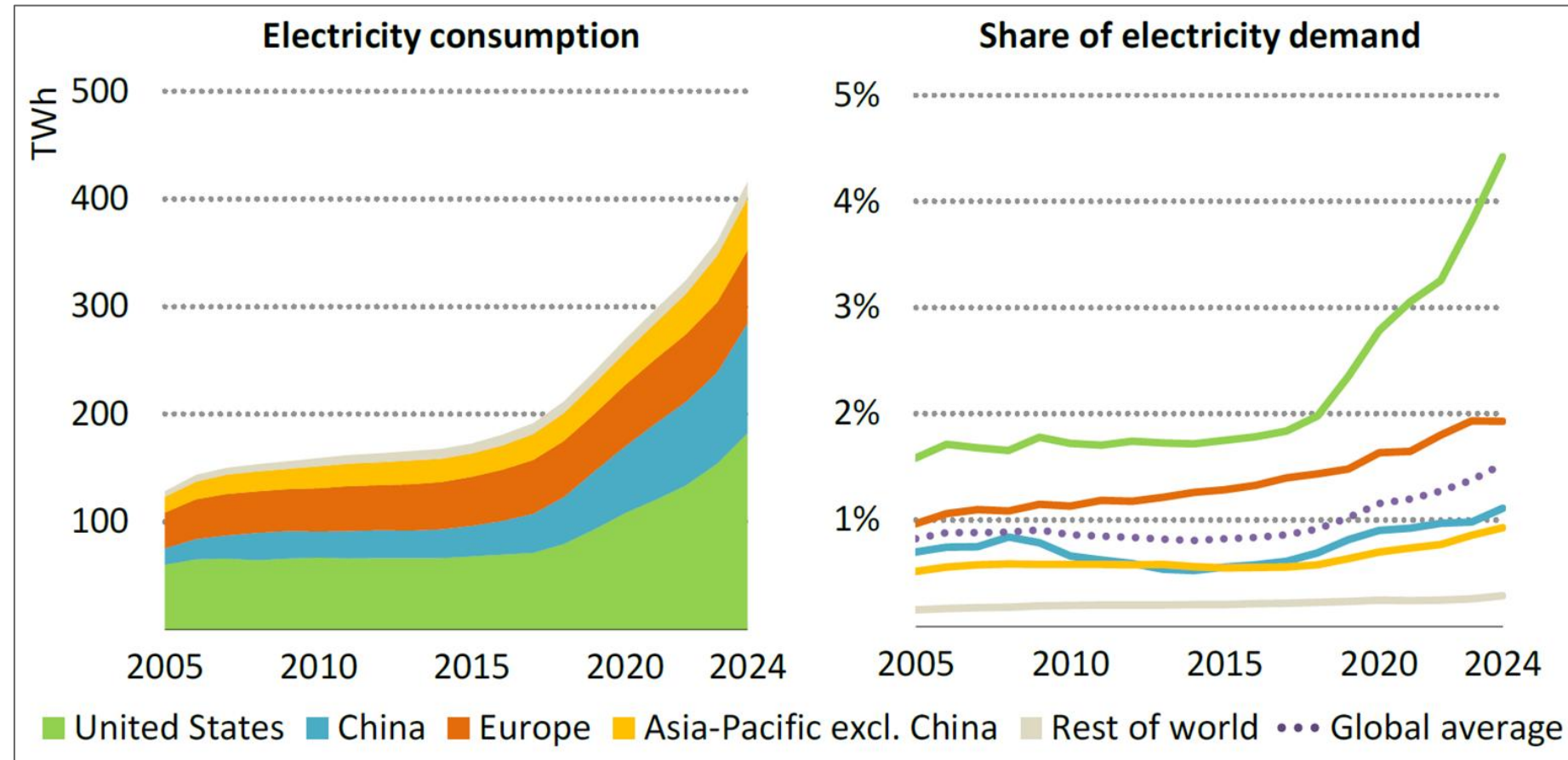
各区域的数据中心用电量 = 根据国际能源署（IEA）的数据，美国领先

Data Center Electricity Consumption by Region – 2005-2024, per IEA



Source: International Energy Agency (IEA), 'Energy and AI' (4/25)

各区域的数据中心用电量 – 2005-2024 年，根据国际能源署（IEA）的数据



Source: International Energy Agency (IEA), 'Energy and AI' (4/25)

- 1 Seem Like Change Happening Faster Than Ever?  
Yes, It Is
- 2 AI User + Usage + CapEx Growth =  
Unprecedented
- 3 AI Model Compute Costs High / Rising + Inference Costs Per Token Falling =  
Performance Converging + Developer Usage Rising
- 4 AI Usage + Cost + Loss Growth =  
Unprecedented
- 5 AI Monetization Threats =  
Rising Competition + Open-Source Momentum + China's Rise
- 6 AI & Physical World Ramps =  
Fast + Data-Driven
- 7 Global Internet User Ramps Powered by AI from Get-Go =  
Growth We Have Not Seen Likes of Before
- 8 AI & Work Evolution =  
Real + Rapid

- 1 变化似乎比以往任何时候都快？是的，的确如此
- 2 AI 用户 + 使用情况 + 资本支出增长 = 前所未有
- 3 AI 模型计算成本高 / 不断上升 + 每个 Token 的推理成本下降 = 性能趋同 + 开发者使用量上升
- 4 AI 使用情况 + 成本 + 损失增长 = 前所未有
- 5 人工智能货币化威胁 = 竞争加剧 + 开源势头 + 中国的崛起
- 6 AI 与物理世界的加速 = 快速 + 数据驱动
- 7 全球互联网用户增长由人工智能驱动 = 前所未有的增长
- 8 人工智能与工作变革 = 真实的 + 迅速的



To understand where AI model economics may be heading, one can look at the mounting tension between capabilities and costs.

Training the most powerful large language models (LLMs) has become one of the most expensive / capital-intensive efforts in human history. As the frontier of performance pushes toward ever-larger parameter counts and more complex architectures, model training costs are rising into the billions of dollars.

Ironically, this race to build the most capable general-purpose models may be accelerating commoditization and driving diminishing returns, as output quality converges across players and differentiation becomes harder to sustain. At the same time, the cost of applying/using these models – known as inference – is falling quickly. Hardware is improving – for example, NVIDIA’s 2024 Blackwell GPU consumes 105,000x less energy per token than its 2014 Kepler GPU predecessor. Couple that with breakthroughs in models’ algorithmic efficiency, and the cost of inference is plummeting.

Inference represents a new cost curve, and – unlike training costs – it’s arcing down, not up. As inference becomes cheaper and more efficient, the competitive pressure amongst LLM providers increases – not on accuracy alone, but also on latency, uptime, and cost-per-token\*. What used to cost dollars can now cost pennies. And what cost pennies may soon cost fractions of a cent.

The implications are still unfolding. For users (and developers), this shift is a gift: dramatically lower unit costs to access powerful AI. And as end-user costs decline, creation of new products and services is flourishing, and user and usage adoption is rising.

For model providers, however, this raises real questions about monetization and profits. Training is expensive, serving is getting cheap, and pricing power is slipping. The business model is in flux. And there are new questions about the one-size-fits-all LLM approach, with smaller, cheaper models trained for custom use cases\*\* now emerging.

Will providers try to build horizontal platforms? Will they dive into specialized applications? Only time will tell. In the short term, it’s hard to ignore that the economics of general-purpose LLMs look like commodity businesses with venture-scale burn.

*\*Cost-per-token = The expense incurred for processing or generating a single token (a word, sub-word, or character) during the operation of a language model. It is a key metric used to evaluate the computational efficiency and cost-effectiveness of deploying AI models, particularly in applications like natural language processing.*  
*\*\*E.g., OpenEvidence*

To understand where AI model economics may be heading, one can look at the mounting tension between capabilities and costs.

训练最强大的大型语言模型 (LLM) 已成为人类历史上最昂贵 / 资本密集的努力之一。随着性能的前沿向着越来越大的参数计数和更复杂的架构推进，模型训练成本正在上升到数十亿美元。

具有讽刺意味的是，这场构建最强大的通用模型竞赛可能会加速商品化并导致收益递减，因为各参与者的输出质量趋同，并且差异化变得更难维持。与此同时，应用 / 使用这些模型的成本 – 被称为推理 – 正在迅速下降。硬件正在改进 – 例如，NVIDIA 的 2024 Blackwell GPU 每 token 的能耗比其 2014 Kepler GPU 前代产品低 105,000 倍。再加上模型算法效率方面的突破，推理成本正在暴跌。

推理代表了一条新的成本曲线，并且 – 与训练成本不同 – 它是向下弯曲，而不是向上。随着推理变得更便宜和更高效，LLM 提供商之间的竞争压力增加 – 不仅仅是在准确性方面，还在延迟、正常运行时间和每次 token 的成本 \* 方面。过去需要花费数美元的现在可能只需要花费几美分。而过去需要花费几美分的可能很快只需要花费几分之一美分。

其影响仍在不断显现。对于用户（和开发者）而言，这种转变是一份礼物：以极低的单位成本访问强大的 AI。并且随着最终用户成本的下降，新产品和服务的创造蓬勃发展，用户和使用采纳率也在上升。

然而，对于模型提供商而言，这引发了关于货币化和利润的实际问题。训练成本高昂，服务成本越来越低，定价权正在下降。商业模式正在变化。而且，关于一刀切的 LLM 方法也出现了新的问题，针对自定义用例训练的更小、更便宜的模型 \*\* 现在正在出现。

供应商会尝试构建横向平台吗？他们会投入到专业应用中吗？只有时间会证明。短期内，很难忽视通用 LLM 的经济效益看起来像是具有风险规模消耗的商品业务。

\* 每次 token 的成本 = 在语言模型运行期间，处理或生成单个 token （单词、子词或字符）所产生的费用。它是评估部署 AI 模型（尤其是在自然语言处理等应用中）的计算效率和成本效益的关键指标。 \*\* 例如，OpenEvidence

***AI Model Compute Costs High / Rising***

+

*Inference Costs Per Token Falling*

=

*Performance Converging + Developer Usage Rising*

AI 模型计算成本高 / 上涨 +

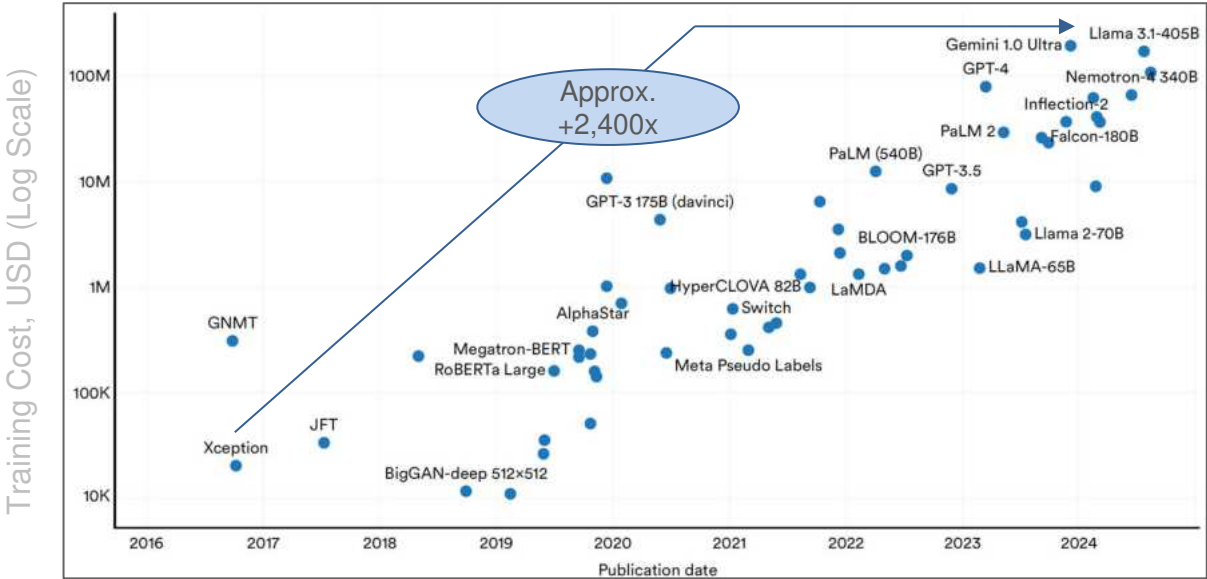
每次 Token 的推断成本下降 =

性能趋同 + 开发者使用量上升



# AI Model Training Compute Costs = ~2,400x Growth Over Eight Years, per Epoch AI & Stanford

Estimated Training Cost of Frontier AI Models – 2016-2024, per Epoch AI & Stanford



Right now, [AI model training costs] \$100 million. There are models in training today that are more like a billion. Right. I think if we go to \$10 or \$100 billion, and I think that will happen in 2025, 2026, maybe 2027...

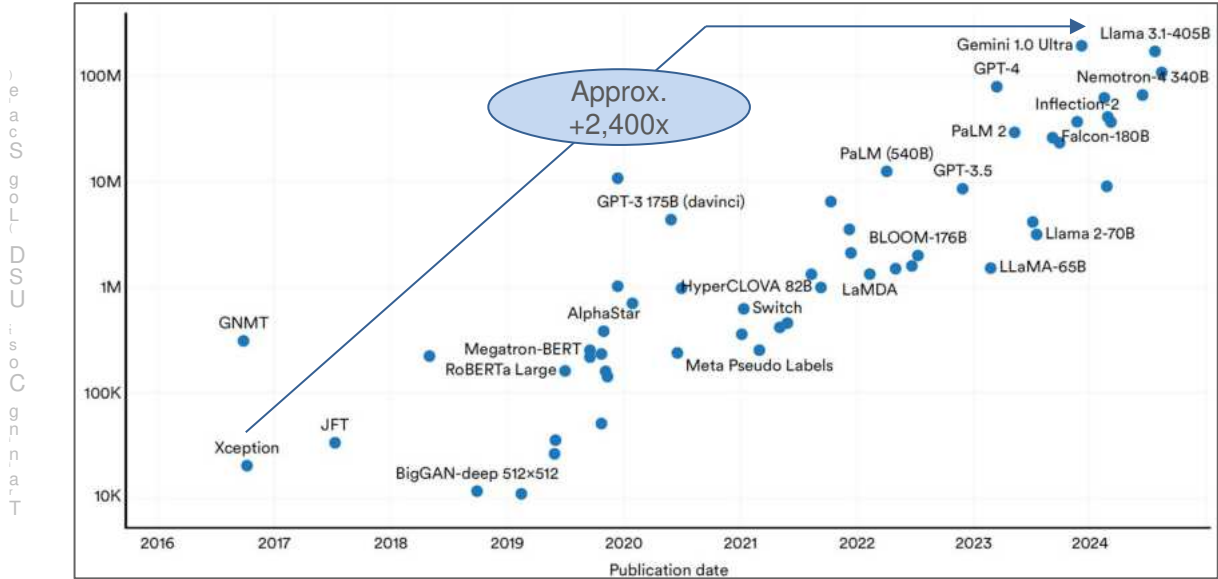
...I think that the training of...\$10 billion models, yeah, could start sometime in 2025.

- Anthropic Co-Founder & CEO Dario Amodei (6/24)

Note: Costs are estimates. Excludes most Chinese models due to lack of reliable cost data. Source: Epoch AI via Nestor Maslej et al., 'The AI Index 2025 Annual Report,' AI Index Steering Committee, Stanford HAI (4/25); In Good Company podcast (6/24)

# AI 模型训练计算成本 = ~2, 八年内增长 400 倍, 每个 Epoch AI & Stanford

前沿 AI 模型训练成本估算 – 2016-2024, 每个 Epoch AI & Stanford



目前, [AI 模型训练成本] 为 1 亿美元。现在正在训练的模型更像是 10 亿美元。没错。我认为如果我们达到 100 亿或 1000 亿美元, 我认为这将会在 2025 年、2026 年, 也许 2027 年发生 .....

..... 我认为 ...100 亿美元模型的训练可能会在 2025 年的某个时候开始。

- Anthropic 联合创始人兼首席执行官 Dario Amodei (6/24)

注意: 成本为估计值。由于缺乏可靠的成本数据, 不包括大多数中国模型。来源: Epoch AI via Nestor Maslej et al., 'The AI Index 2025 Annual Report,' AI Index Steering Committee, Stanford HAI (4/25); In Good Company podcast (6/24)

*AI Model Compute Costs High / Rising*  
+  
**Inference Costs Per Token Falling**  
=  
*Performance Converging + Developer Usage Rising*

AI 模型每算成本高昂性能提升  
+  
推理成本下降开发者使用量上升  
=

---

To understand the trajectory of AI compute, it helps to revisit an idea from the early days of PC software. ‘Software is a gas...it expands to fill its container,’ said Nathan Myhrvold, then CTO of Microsoft in 1997. AI is proving no different. As models get better, usage increases – and as usage increases, so does demand for compute. We’re seeing it across every layer: more queries, more models, more tokens per task. The appetite for AI isn't slowing down. It’s growing into every available resource – just like software did in the age of desktop and cloud.

But infrastructure is not just standing still. In fact, it's advancing faster than almost any other layer in the stack, and at unprecedented rates. As noted on page 136, NVIDIA’s 2024 Blackwell GPU uses 105,000 times less energy to generate tokens than its 2014 Kepler predecessor. It’s a staggering leap, and it tells a deeper story – not just of cost reduction, but of architectural and materials innovation that is reshaping what’s possible at the hardware level.

These improvements in hardware efficiency are critical to offset the strain of increasing AI and internet usage on our grid. So far, though, they have not been enough. This trend aligns with Jevons Paradox, first proposed back in 1865\* – that technological advancements that improve resource efficiency actually lead to increased overall usage of those resources. This is driving new focus on expanding energy production capacity – and new questions about the grid’s ability to manage.

Yet again, we see this as one of the perpetual ‘a-ha’s’ of technology: costs fall, performance rises, and usage grows, all in tandem. This trend is repeating itself with AI.

*\*British economist William Stanley Jevons first observed this phenomenon in 19<sup>th</sup>-century Britain, where he noticed that improvements in the efficiency of coal-powered steam engines were not reducing coal consumption but rather increasing it. In his book The Coal Question, he noted 'It is wholly a confusion of ideas to suppose that the economical use of fuel is equivalent to diminished consumption. The very contrary is the truth.'*

要理解人工智能计算的轨迹，回顾一下 PC 软件早期的想法会有所帮助。“ 软件是一种气体 …… 它会膨胀以填充其容器 ”，微软当时的首席技术官 Nathan Myhrvold 在 1997 年说道。人工智能正在证明也没有什么不同。随着模型变得更好，使用量也会增加 —— 并且随着使用量的增加，对计算的需求也会增加。我们正在每一层都看到这种情况：更多的查询、更多的模型、每个任务更多的 token。对人工智能的需求并没有放缓。它正在扩展到每一个可用的资源中 —— 就像软件在桌面和云时代所做的那样。

但基础设施并没有停滞不前。事实上，它的发展速度比堆栈中的几乎任何其他层都要快，并且速度空前。正如第 136 页指出的那样，NVIDIA 的 2024 Blackwell GPU 产生 token 所消耗的能量比其 2014 年的 Kepler 前身少 105,000 倍。这是一个惊人的飞跃，它讲述了一个更深层次的故事 —— 不仅是成本降低，而且是架构和材料创新，这些创新正在重塑硬件层面上的可能性。

硬件效率的这些改进对于抵消日益增长的 AI 和互联网使用对我们电网的压力至关重要。但到目前为止，这些改进还不够。这一趋势与杰文斯悖论相符，该悖论最早于 1865 年提出 \* – 即提高资源效率的技术进步实际上会导致总体使用量增加那些资源。这正在推动人们重新关注扩大能源生产能力 – 以及关于电网管理能力的新问题。

再一次，我们将其视为技术永恒的 “a- 哈 ” 之一：成本下降，性能上升，使用量增长，所有这些都是同步的。这种趋势正在 AI 中重演。

*\*British economist William Stanley Jevons first observed this phenomenon in 19<sup>th</sup>-century Britain, where he noticed that improvements in the efficiency of coal-powered steam engines were not reducing coal consumption but rather increasing it. In his book The Coal Question, he noted 'It is wholly a confusion of ideas to suppose that the economical use of fuel is equivalent to diminished consumption. The very contrary is the truth.'*

What are tokens and how to count them?

Updated over 3 months ago

What are tokens?

Tokens can be thought of as pieces of words. Before the API processes the request, the input is broken down into tokens. These tokens are not cut up exactly where the words start or end - tokens can include trailing spaces and even sub-words. Here are some helpful rules of thumb for understanding tokens in terms of lengths:

- 1 token ~= 4 chars in English
- 1 token ~= ¾ words
- 100 tokens ~= 75 words

Or

- 1-2 sentence ~= 30 tokens
- 1 paragraph ~= 100 tokens
- 1,500 words ~= 2048 tokens

To get additional context on how tokens stack up, consider this:

- Wayne Gretzky's quote "You miss 100% of the shots you don't take" contains 11 tokens.
- OpenAI's charter contains 476 tokens.
- The transcript of the US *Declaration of Independence* contains 1,695 tokens.

\*Assumes that the average ChatGPT interaction consumes 200 total tokens (input + output), or 150 words. Thus, 1MM tokens equates to roughly 5,000 ChatGPT responses. Source: OpenAI (1/25)

Additional context: 1MM tokens = ~750,000 words...roughly

- 3,500 pages of a standard book (12-point font, double-spaced)
- 5,000 ChatGPT responses\*

What are tokens and how to count them?

Updated over 3 months ago

What are tokens?

Tokens can be thought of as pieces of words. Before the API processes the request, the input is broken down into tokens. These tokens are not cut up exactly where the words start or end - tokens can include trailing spaces and even sub-words. Here are some helpful rules of thumb for understanding tokens in terms of lengths:

- 1 token ~= 4 chars in English
- 1 token ~= ¾ words
- 100 tokens ~= 75 words

Or

- 1-2 sentence ~= 30 tokens
- 1 paragraph ~= 100 tokens
- 1,500 words ~= 2048 tokens

To get additional context on how tokens stack up, consider this:

- Wayne Gretzky's quote "You miss 100% of the shots you don't take" contains 11 tokens.
- OpenAI's charter contains 476 tokens.
- The transcript of the US *Declaration of Independence* contains 1,695 tokens.

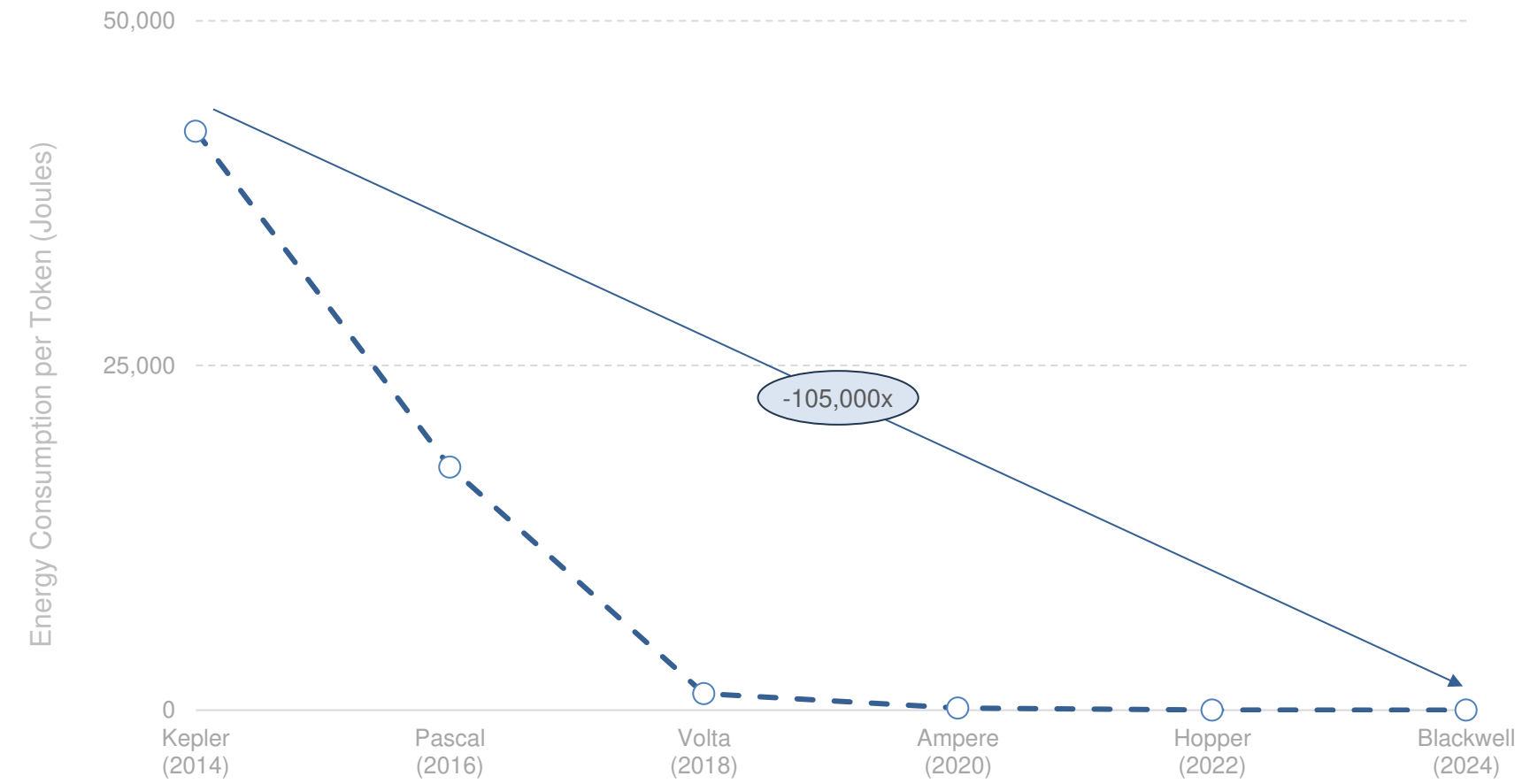
\*Assumes that the average ChatGPT interaction consumes 200 total tokens (input + output), or 150 words. Thus, 1MM tokens equates to roughly 5,000 ChatGPT responses. Source: OpenAI (1/25)

补充说明：1MM tokens =~750,000 词 …… 大致

- 3, 500 页标准书籍（12 号字体，双倍行距）
- 5,000 个 ChatGPT 回复 \*

# AI Inference Costs – NVIDIA GPUs = -105,000x Decline in Energy Required to Generate Token Over Ten Years

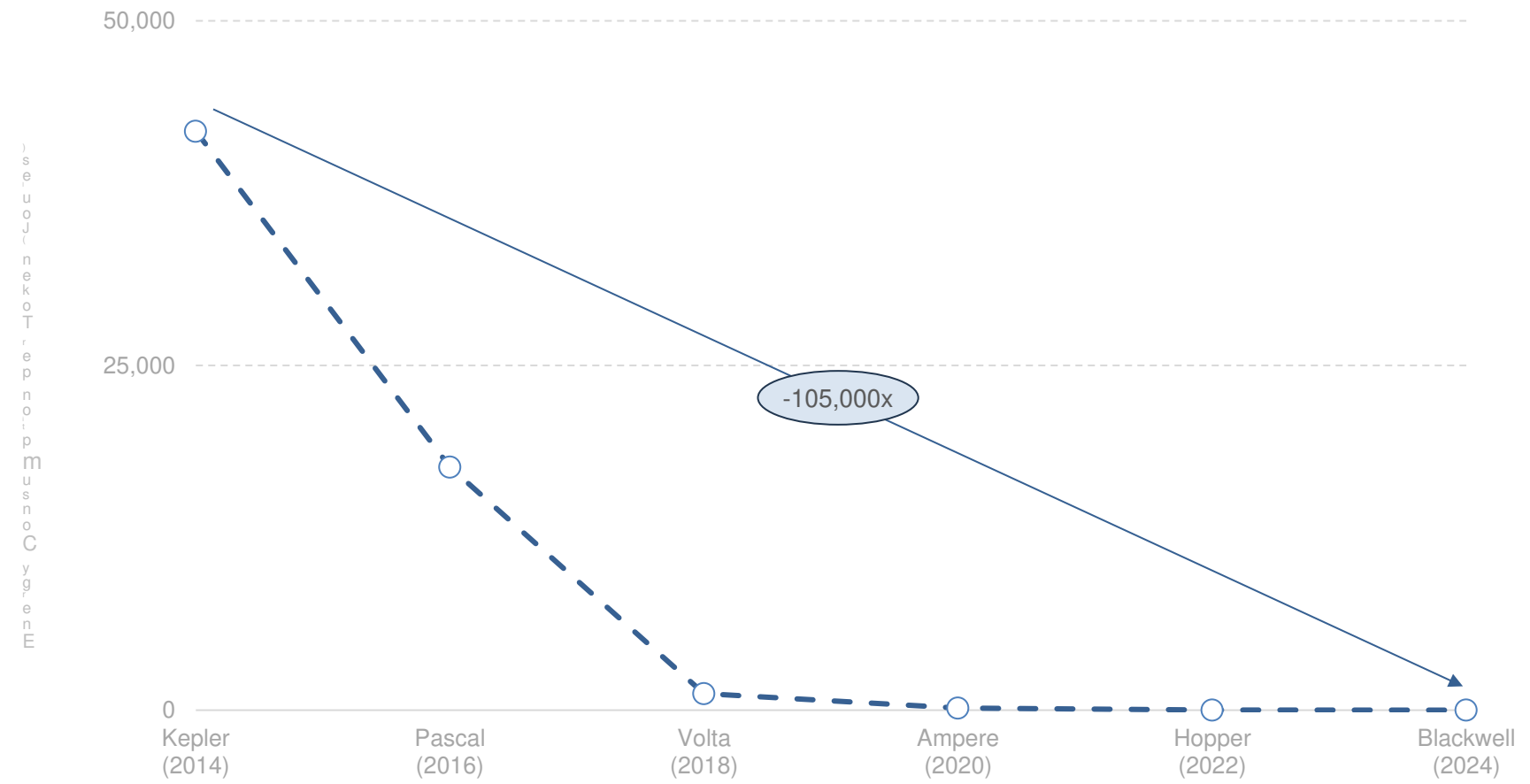
Energy Required per LLM Token (Joules), NVIDIA GPUs – 2014-2024, per NVIDIA



Note: Kepler released in 2012. NVIDIA materials mark performance threshold shown above for Kepler as of 2014. Source: NVIDIA Company Overview (2/25)

# AI Inference Costs – NVIDIA GPUs = -105,000x Decline in Energy Required to Generate Token Over Ten Years

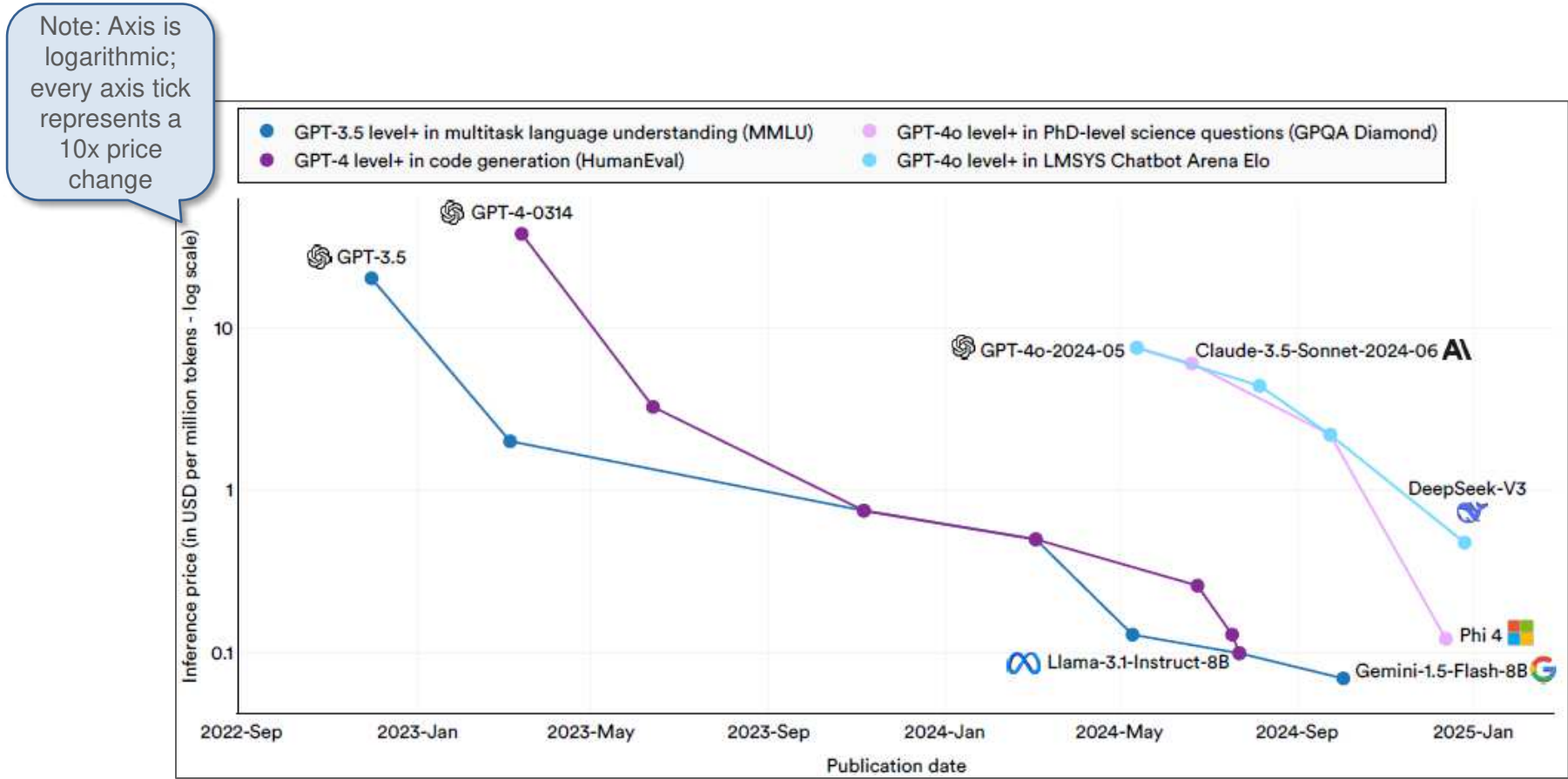
Energy Required per LLM Token (Joules), NVIDIA GPUs – 2014-2024, per NVIDIA



Note: Kepler released in 2012. NVIDIA 材料将 2014 年 Kepler 的性能阈值标记如上。来源: NVIDIA 公司概览 (2/25)

# AI Inference Costs – Serving Models = 99.7% Lower Over Two Years, per Stanford HAI

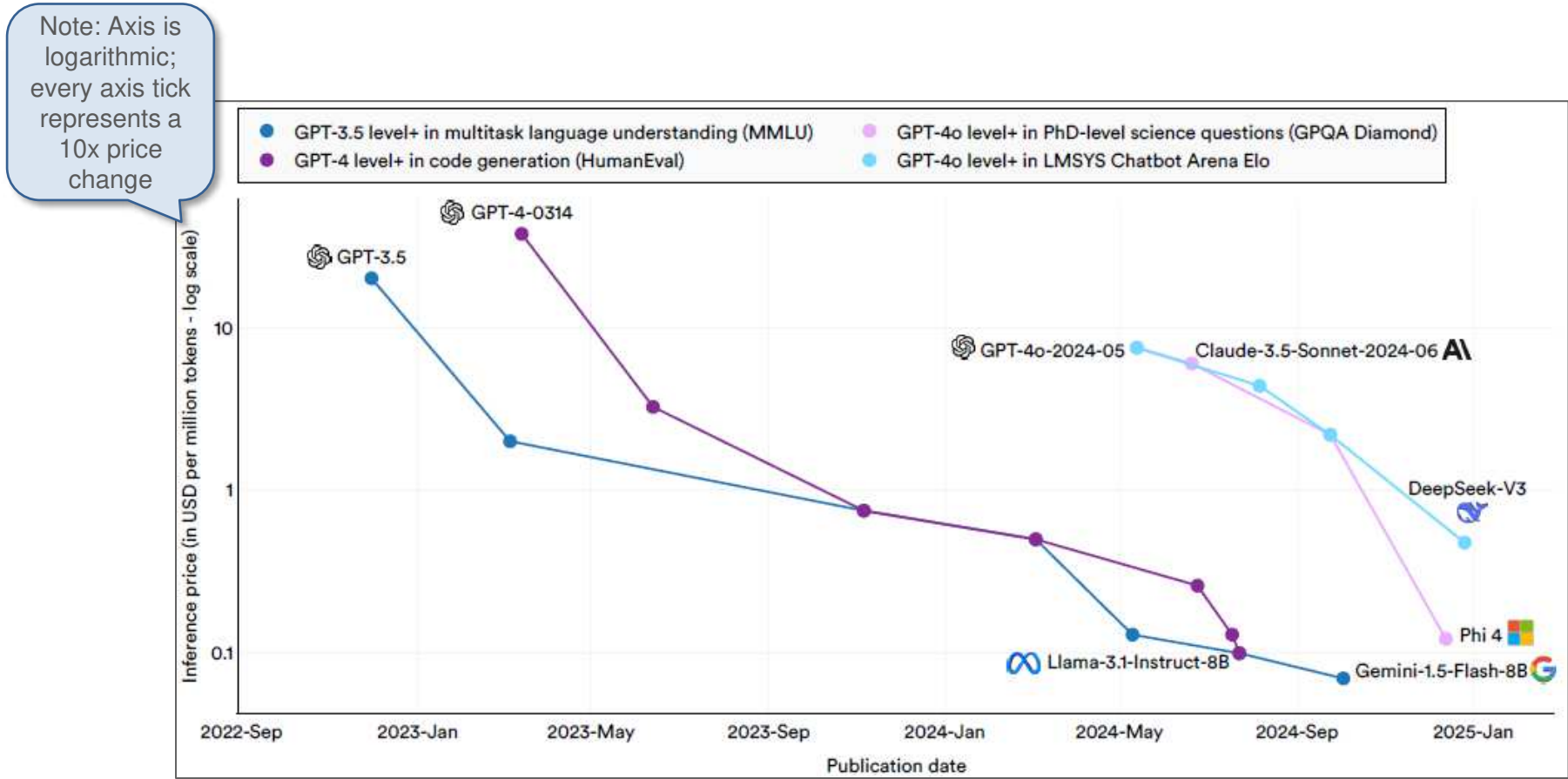
AI Inference Price for Customers (per 1 Million Tokens) – 11/22-12/24, per Stanford HAI



Source: Nestor Maslej et al., 'The AI Index 2025 Annual Report,' AI Index Steering Committee, Stanford HAI (4/25)

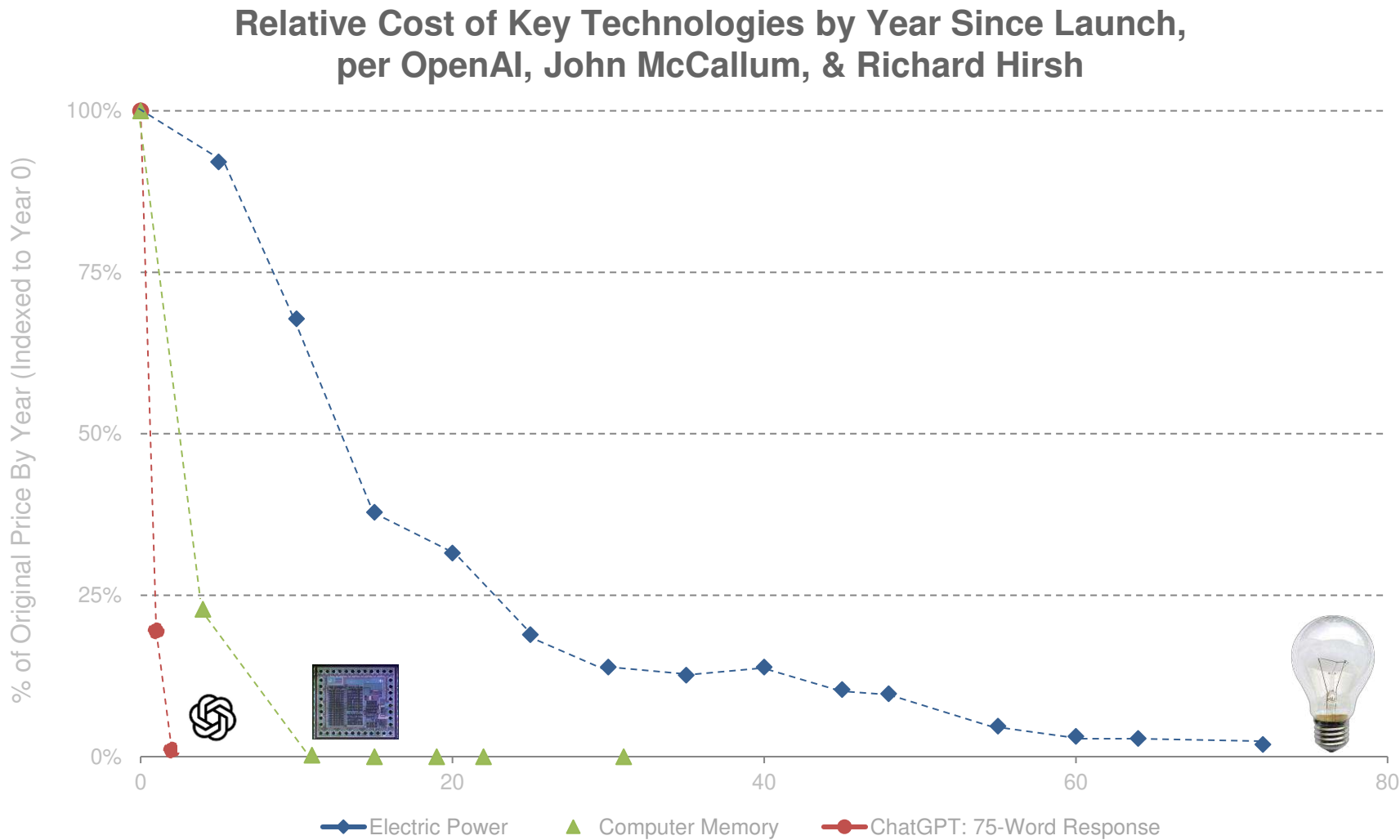
# AI Inference Costs – Serving Models = 99.7% Lower Over Two Years, per Stanford HAI

面向客户的 AI 推理价格（每 100 万个 tokens） – 11/22-12/24, 来源: Stanford HAI



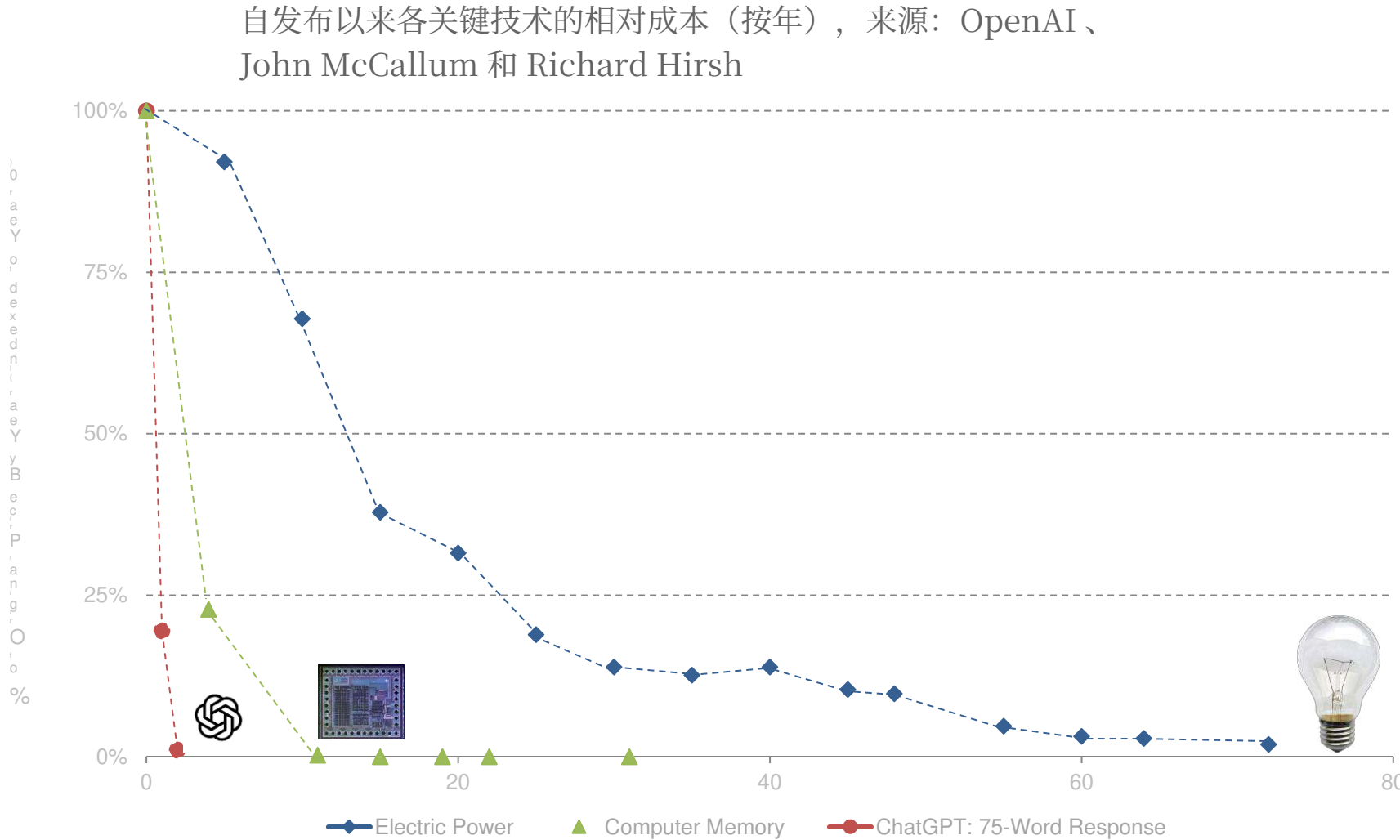
来源: Nestor Maslej 等人, 《AI Index 2025 年度报告》, AI Index Steering Committee, Stanford HAI (4/25)

# AI Cost Efficiency Gains = Happening Faster vs. Prior Technologies



Note: Price change in consumer goods and services in the United States is measured as the percentage change since 1997. Data is based on the consumer price index (CPI) for national average urban consumer prices. Per OpenAI, 100 AI 'tokens' generates approximately 75 words in a large language model response; data shown indexes to this number of tokens. 'Year 0' is not necessarily the year that the technology was introduced, but rather the first year of available data.  
Source: Electricity Costs – Technology and Transformation in the American Electric Utility Industry, Richard Hirsh (1989); Computer Memory Storage Costs – John C. McCallum, with data aggregated from 72 primary sources and historical company sales documents; OpenAI, Wikimedia Commons

# AI Cost Efficiency Gains = Happening Faster vs. Prior Technologies



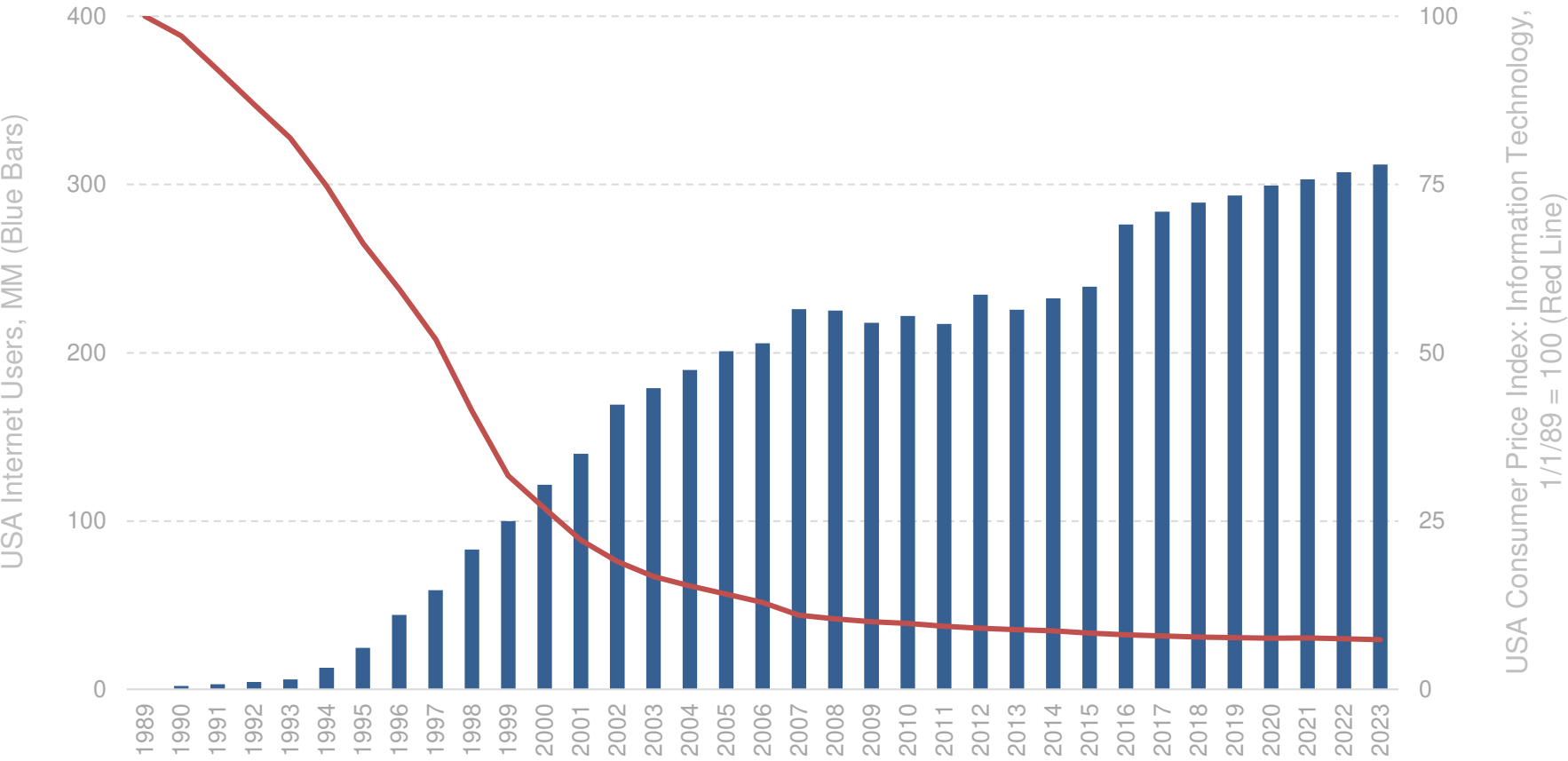
Note: Price change in consumer goods and services in the United States is measured as the percentage change since 1997. Data is based on the consumer price index (CPI) for national average urban consumer prices. Per OpenAI, 100 AI 'tokens' generates approximately 75 words in a large language model response; data shown indexes to this number of tokens. 'Year 0' is not necessarily the year that the technology was introduced, but rather the first year of available data.  
Source: Electricity Costs – Technology and Transformation in the American Electric Utility Industry, Richard Hirsh (1989); Computer Memory Storage Costs – John C. McCallum, with data aggregated from 72 primary sources and historical company sales documents; OpenAI, Wikimedia Commons



Tech’s Perpetual A-Ha = Declining Costs + Improving Performance → Rising Adoption...

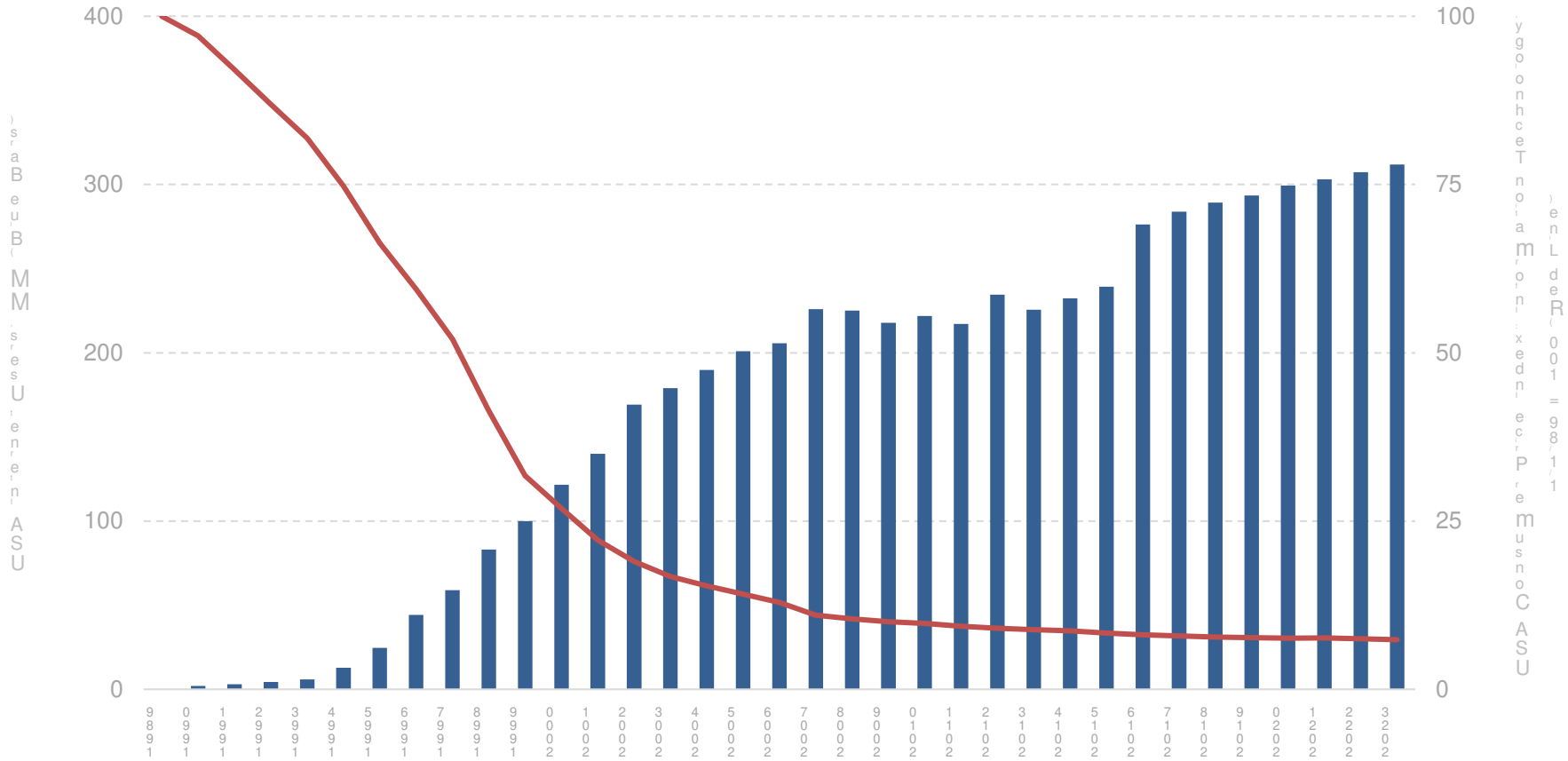
科技的永恒灵感 -Ha = 成本下降 + 性能提升 → 采用率上升 …

USA Internet Users (MM) vs. Relative IT Cost – 1989-2023, per FRED & ITU



Note: FRED data shows 'Consumer Price Index for All Urban Consumers: Information Technology, Hardware and Services in U.S. City Average.' Source: USA Federal Reserve Bank of St. Louis (FRED), International Telecommunications Union (via World Bank) (4/25)

USA Internet Users (MM) vs. Relative IT Cost – 1989-2023, per FRED & ITU

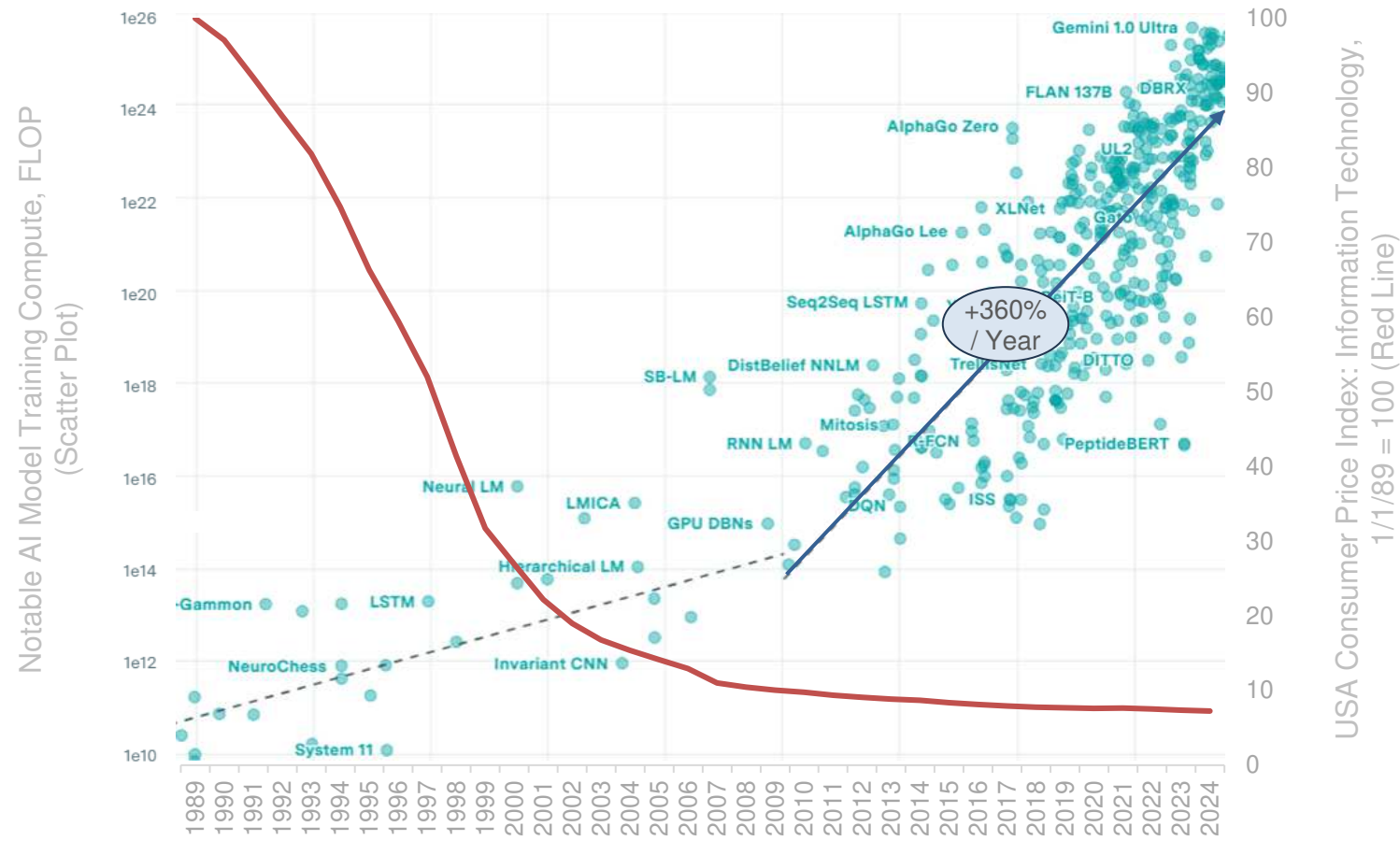


注意：FRED 数据显示 “美国所有城市消费者的消费价格指数：信息技术、硬件和服务”。平均值。”来源：美国圣路易斯联邦储备银行 (FRED)，国际电信联盟（通过世界银行）(4/25)



...Tech's Perpetual A-Ha =  
Prices Fall + Performance Rises

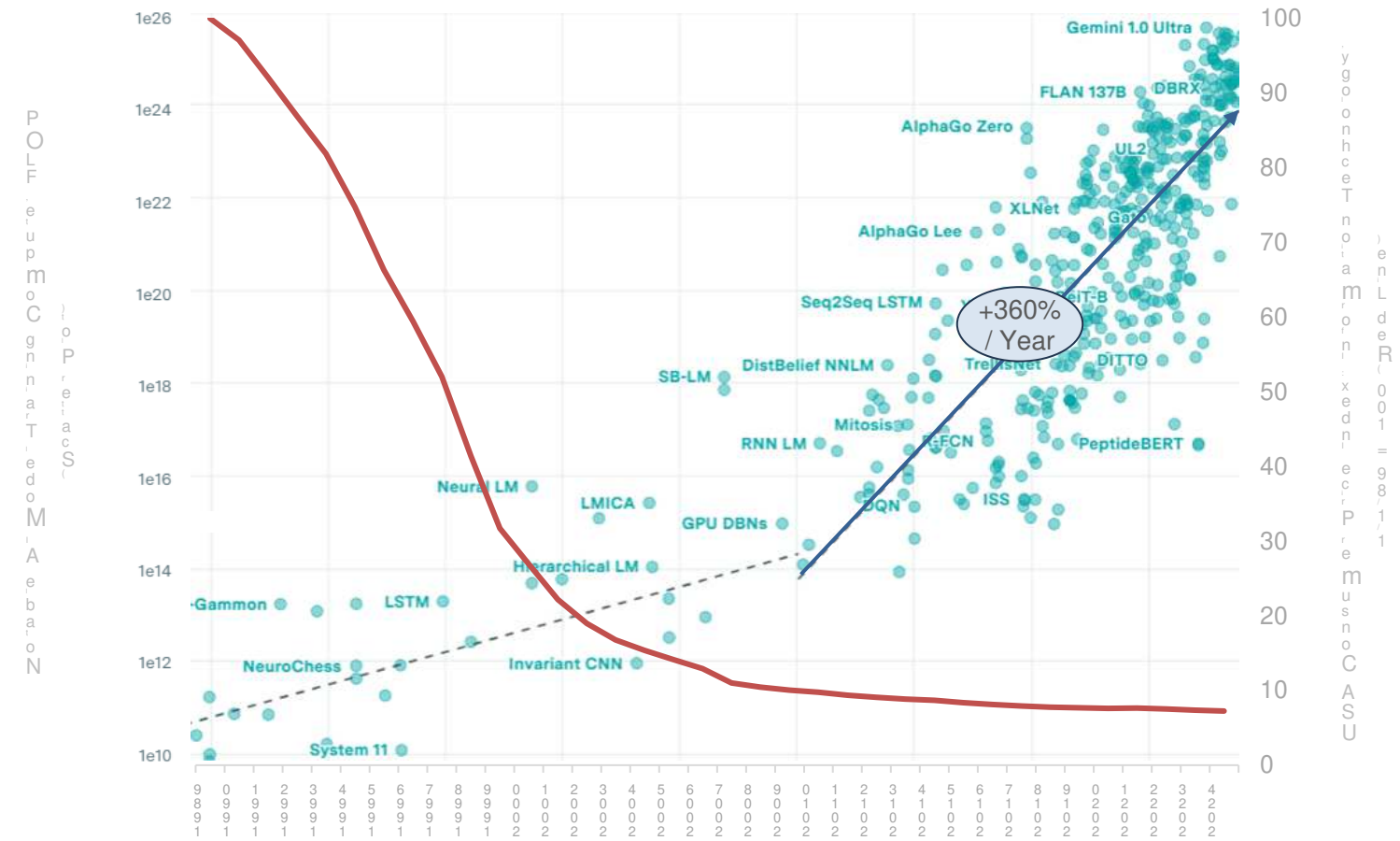
AI Model Training Compute (FLOP) vs. Relative IT Cost – 1989-2024, per Epoch AI & FRED



Note: A petaFLOP/s-day represents the total computational work performed by a system operating at 1 petaFLOP/s ( $10^{15}$  floating-point operations per second) for 24 hours, equivalent to approximately  $8.64 \times 10^{19}$  operations. This metric is commonly used to quantify the compute required for large-scale tasks like training machine learning models. FRED data shows 'Consumer Price Index for All Urban Consumers: Information Technology, Hardware and Services in U.S. City Average.' Note that, while training compute is not a direct measurement of model performance, it is typically closely correlated with performance. Source: USA Federal Reserve Bank of St. Louis (FRED); Epoch AI (5/25)

...Tech's Perpetual A-Ha = 价格下  
降 + 性能提升

AI Model Training Compute (FLOP) vs. Relative IT Cost – 1989-2024, per Epoch AI & FRED



注意：petaFLOP/s-day 表示一个系统以 1 petaFLOP/s ( $10^{15}$  floating-point 每秒运算次数) 运行 24 小时所执行的总计算工作量，相当于大约  $8.64 \times 10^{19}$  次运算。此指标通常用于量化大型 -scale 任务（如训练机器学习模型）所需的计算量。FRED 数据显示 '所有城市消费者的消费者价格指数：美国城市平均的信息技术、硬件和服务。' 请注意，虽然训练计算不是模型性能的直接衡量标准，但它通常与性能密切相关。来源：美国圣路易斯联邦储备银行 (FRED); Epoch AI (5/25)

*AI Model Compute Costs High / Rising*  
+  
*Inference Costs Per Token Falling*  
=  
**Performance Converging** + *Developer Usage Rising*

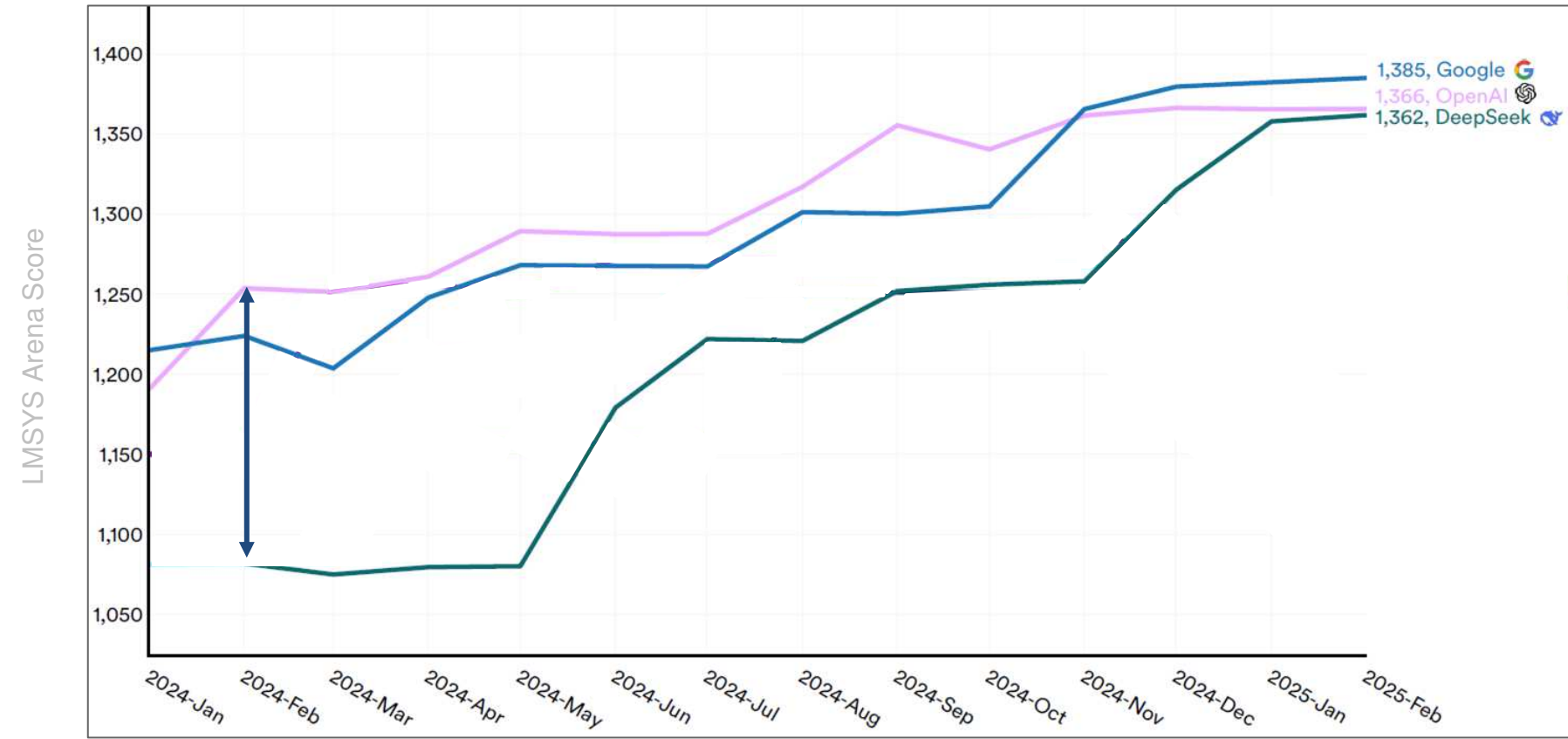
AI 模型每算成本高 / 性能提升  
+  
推理成本下降 / 开发者使用量上升  
=

---

# AI Model Performance = Converging Rapidly, per Stanford HAI

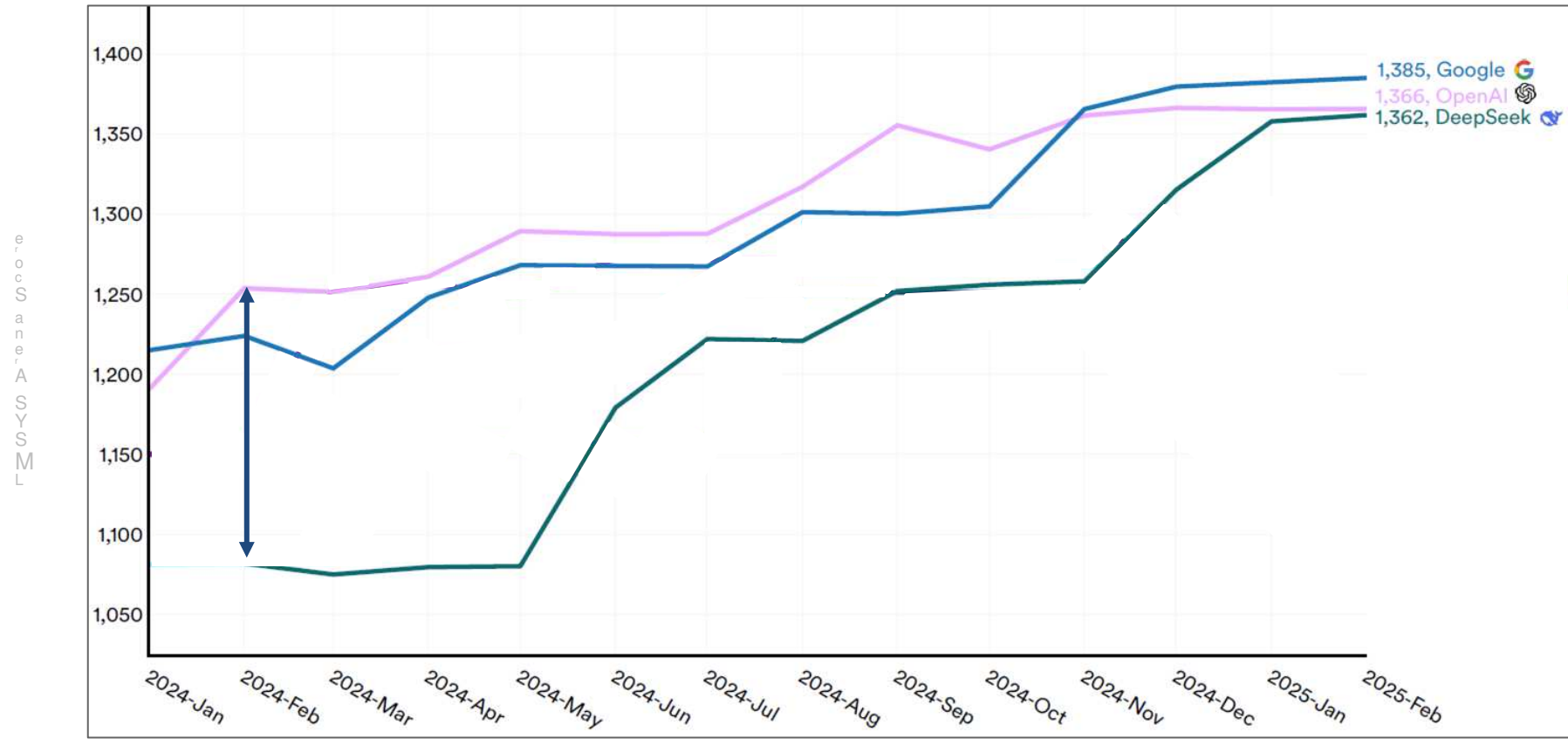
根据斯坦福 HAI，人工智能模型性能 = 正在快速收敛

Performance of Top AI Models on LMSYS Chatbot Arena – 1/24-2/25, per Stanford HAI



Note: The LMSYS Chatbot Arena is a public website where people compare two AI chatbots by asking them the same question and voting on which answer is better. The results help rank how well different language models perform based on human judgment. Only the highest-scoring model in any given month is shown in this comparison. Source: Nestor Maslej et al., "The AI Index 2025 Annual Report," AI Index Steering Committee, Stanford HAI (4/25)

Performance of Top AI Models on LMSYS Chatbot Arena – 1/24-2/25, per Stanford HAI



注：LMSYS Chatbot Arena 是一个公共网站，人们可以在该网站上通过向两个 AI 聊天机器人提出相同的问题并投票选出哪个答案更好来比较它们。结果有助于根据人类判断对不同语言模型的性能进行排名。在此比较中，仅显示任何给定月份中得分最高的模型。来源：Nestor Maslej et al., "The AI Index 2025 Annual Report," AI Index Steering Committee, Stanford HAI (4/25)

*AI Model Compute Costs High / Rising*  
+  
*Inference Costs Per Token Falling*  
=  
*Performance Converging + Developer Usage Rising*

AI 模型每算 token 成本高 / 性能提升  
+  
推理成本下降 = 开发者使用量上升

To understand the surge in AI developer activity, it’s instructive to look at the extraordinary drop in inference costs and the growing accessibility of capable models.

Between 2022 and 2024, the cost-per-token to run language models fell by an estimated 99.7% – a decline driven by massive improvements in both hardware and algorithmic efficiency. What was once prohibitively expensive for all but the largest companies is now within reach of solo developers, independent app builders, researchers on a laptop, and mom-and-pop shop employees.

The cost collapse has made experimentation cheap, iteration fast, and productization feasible for virtually anyone with an idea.

At the same time, performance convergence is shifting the calculus on model selection. The gap between the top-performing frontier models and smaller, more efficient alternatives is narrowing.

For many use cases – summarization, classification, extraction, or routing – the difference in real-world performance is negligible. Developers are discovering they no longer need to pay a premium for a top-tier model to get reliable outputs. Instead, they can run cheaper models locally or via lower-cost API providers and achieve functionally similar results, especially when fine-tuned on task-specific data.

This shift is weakening the pricing leverage of model incumbents and leveling the playing field for AI development...

为了理解人工智能开发者活动的激增，有必要了解推理成本的急剧下降以及强大模型日益普及的情况。

2022 年至 2024 年间，运行语言模型的每 token 成本估计下降了 99.7%–，这一下降是由硬件和算法效率的大幅提高推动的。曾经只有最大的公司才能承受的高昂成本，现在个人开发者、独立应用构建者、在笔记本电脑上进行研究的研究人员以及夫妻店员工都可以承受了。

成本的崩溃使得几乎任何有想法的人都可以进行廉价的实验、快速的迭代和可行的产品化。

与此同时，性能收敛正在改变模型选择的计算方式。表现最佳的前沿模型与更小、更高效的替代方案之间的差距正在缩小。

对于许多用例 – 摘要、分类、提取或路由 –，现实世界中的性能差异可以忽略不计。开发者们发现，他们不再需要为顶级模型支付高额费用才能获得可靠的输出。相反，他们可以在本地或通过低成本的 API 提供商运行更便宜的模型，并获得功能相似的结果，尤其是在针对特定任务的数据进行微调时。

这种转变削弱了现有模型在定价方面的影响力并为人工智能开发创造了公平的竞争环境 ……

...At the platform level, a proliferation of foundation models has created a new kind of flexibility. Developers can now choose between dozens of models – OpenAI’s ChatGPT, Meta’s Llama, Mistral’s Mixtral, Anthropic’s Claude, Google’s Gemini, Microsoft’s Phi, and others – each of which excels in different domains. Some are optimized for reasoning, others for speed or code generation. The result is a move away from vendor lock-in.

Instead of consolidating under a single provider who can gate access or raise prices, developers are distributing their efforts across multiple ecosystems. This plurality of options is empowering a new wave of builders to choose the best-fit model for their technical or financial needs.

What’s emerging is a flywheel of developer-led infrastructure growth. As more developers build AI-native apps, they also create tools, wrappers and libraries that make it easier for others to follow. New front-end frameworks, embedding pipelines, model routers, vector databases, and serving layers are multiplying at an accelerating rate.

Each wave of developer activity reduces the friction for the next, compressing the time from idea to prototype and from prototype to product. In the process, the barrier to building with AI is collapsing – not just in cost, but in complexity. This is no longer just a platform shift. It’s an explosion of creativity.

Technology history has shown – as memorialized by then-Microsoft President Steve Ballmer’s repeat *Developers! Developers! Developers...* at a 2000 Microsoft Developers Conference ([link](#)) – the platform that gets the most consistent developer user and usage momentum – and can scale and steadily improve – wins.

… 在平台层面，基础模型的激增创造了一种新的灵活性。开发人员现在可以在数十种模型之间进行选择 – OpenAI 的 ChatGPT、Meta 的 Llama、Mistral 的 Mixtral、Anthropic 的 Claude、Google 的 Gemini、Microsoft 的 Phi 等等 – 每种模型都擅长不同的领域。有些针对推理进行了优化，有些则针对速度或代码生成。其结果是摆脱了供应商锁定。

开发人员没有可以在可以限制访问或提高价格的单一提供商下进行整合，而是将他们的努力分散在多个生态系统中。这种选择的多样性正在赋能新一代构建者，使其能够为其技术或财务需求选择最合适的模型。

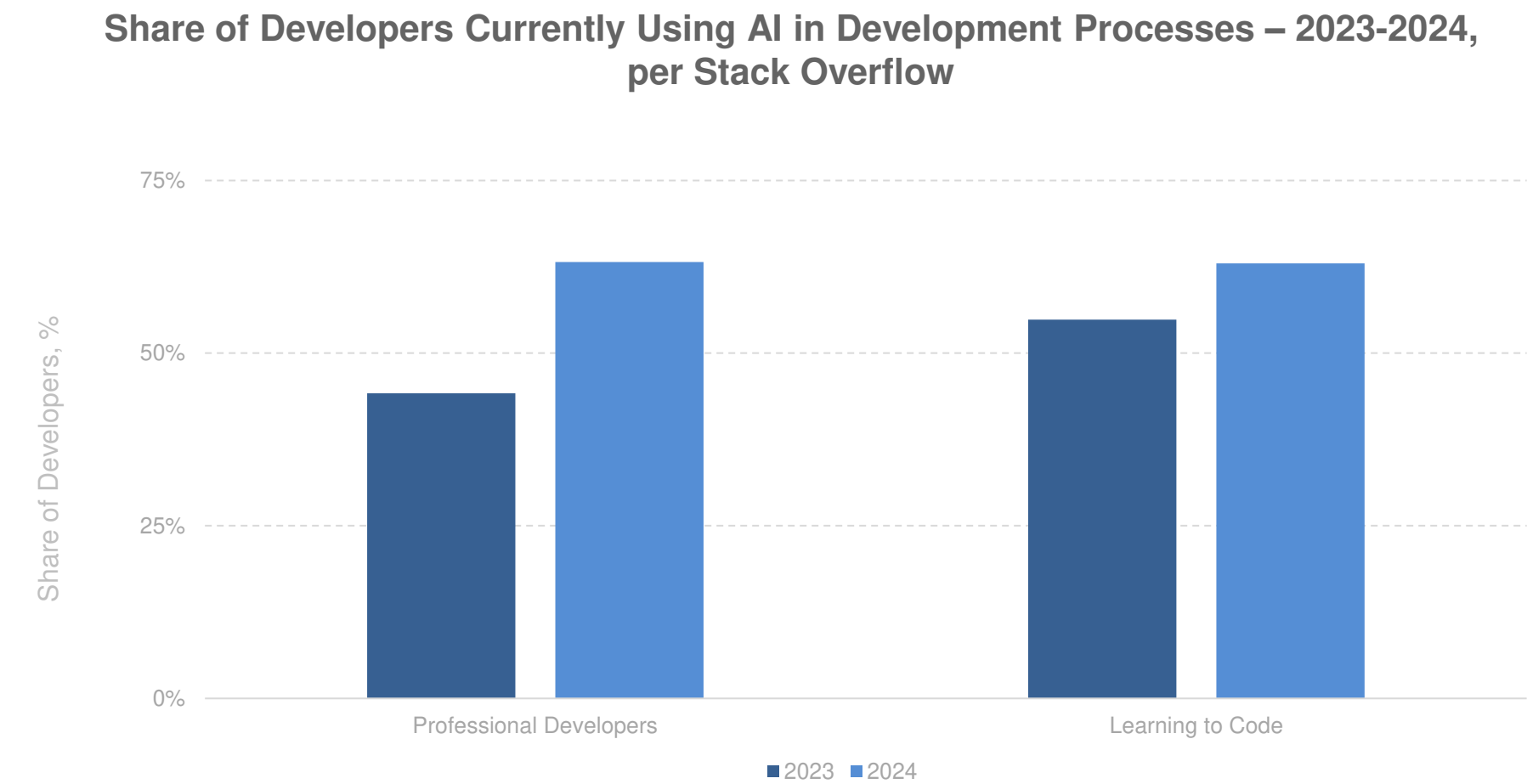
正在出现的是一个由开发者主导的基础设施增长的飞轮。随着越来越多的开发人员构建 AI 原生应用程序，他们还创建了工具、包装器和库，使其他人更容易效仿。新的前端框架、嵌入管道、模型路由器、向量数据库和服务层正在以加速的速度成倍增加。

每一波开发者活动都会减少下一波活动的阻力，从而缩短从想法到原型以及从原型到产品的时间。在此过程中，使用人工智能的障碍正在崩溃 – 不仅在成本方面，而且在复杂性方面。这不再仅仅是一个平台转变。这是一场创造力的爆发。

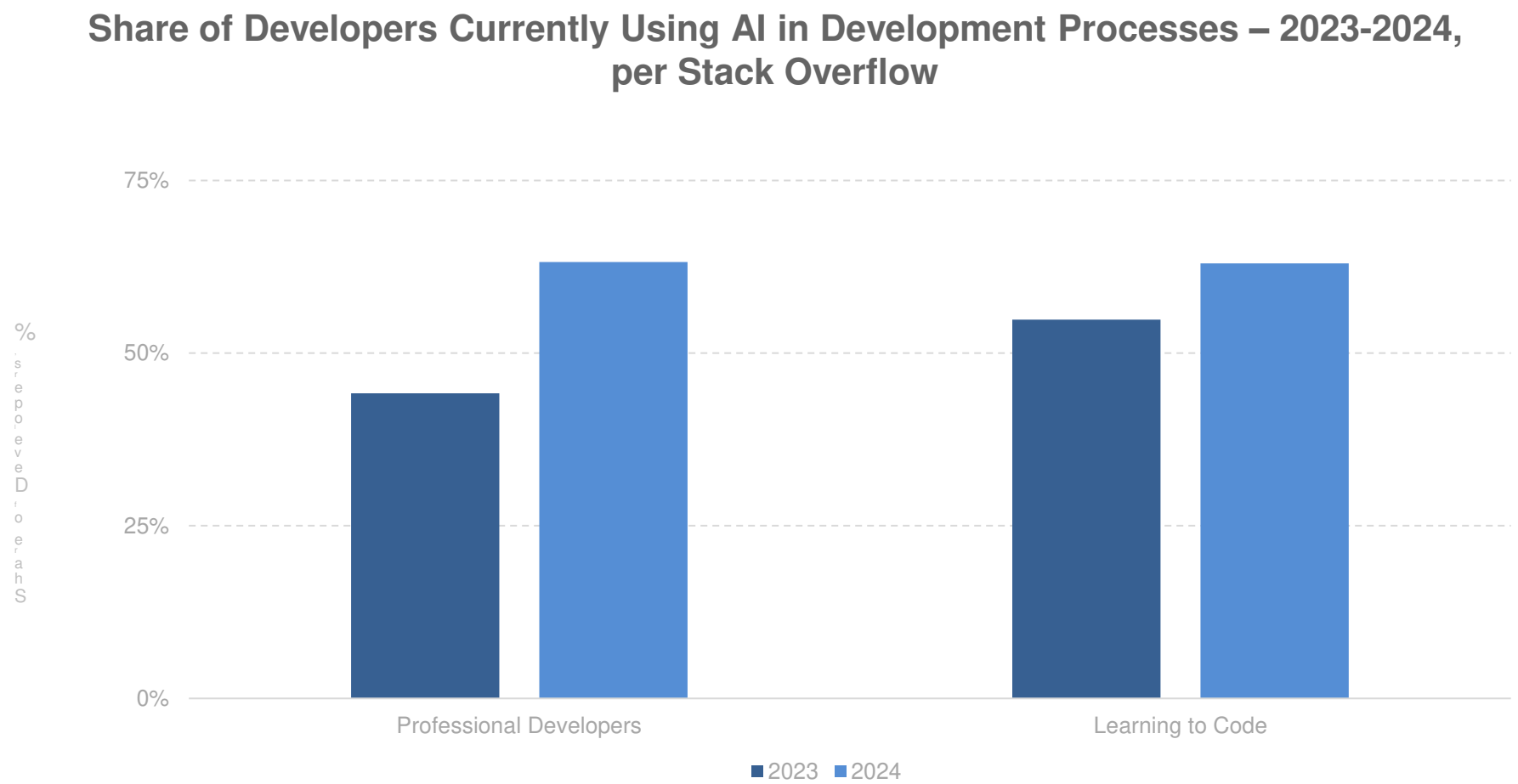
技术历史表明 – 正如时任微软总裁史蒂夫 · 鲍尔默在 2000 年微软开发者大会上重复开发者！开发者！开发者 …… （链接）所纪念的那样 – 获得最持续的开发者用户和使用势头的平台 – 并且可以扩展和稳步改进的平台 – 获胜。

The AI Developer Next Door

隔壁的人工智能开发者



Note: 2023 N=89,184; 2024 N=65,437. Respondents are global. Source: Stack Overflow Developer Surveys (5/23 & 5/24-6/24)



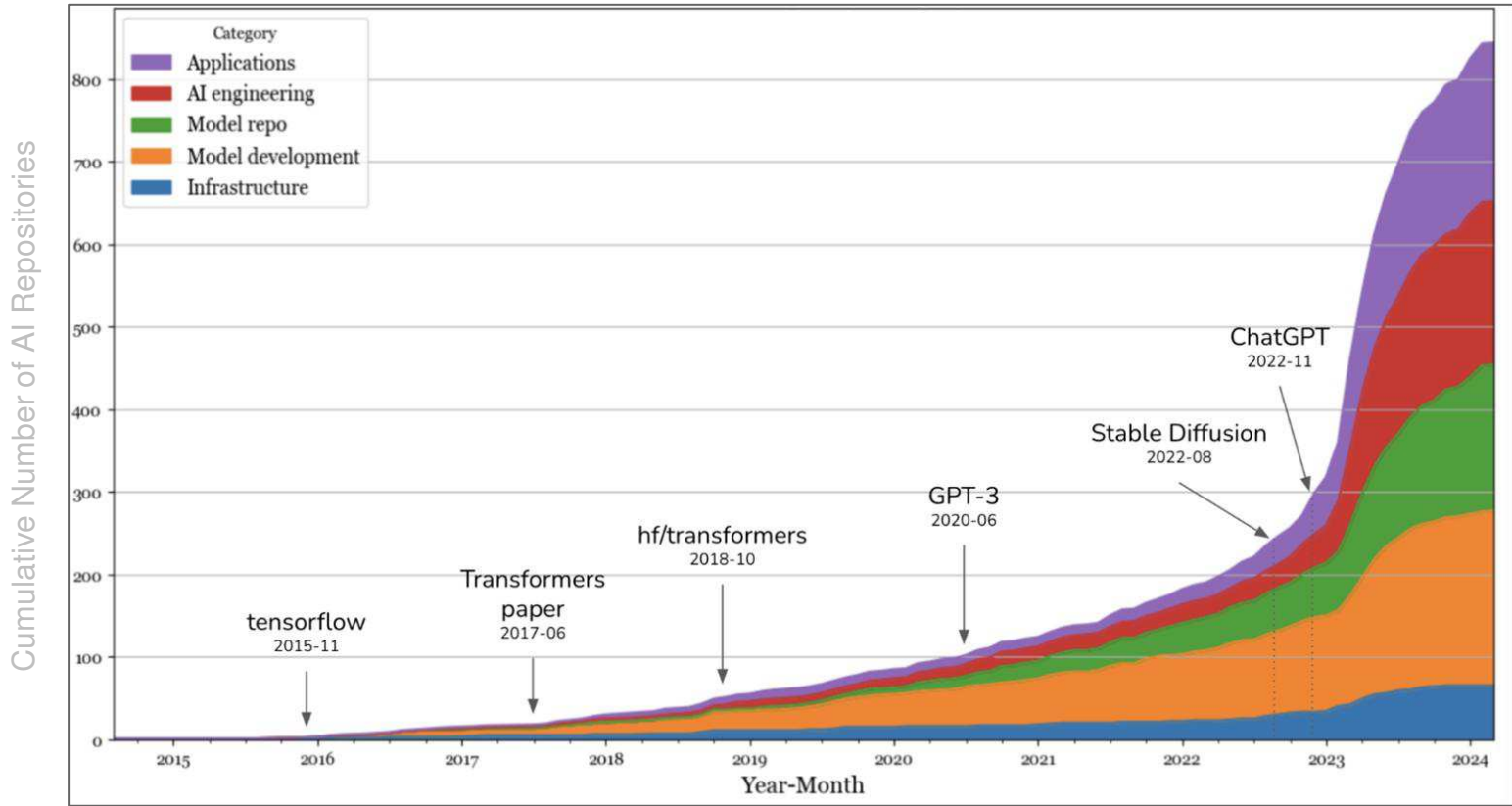
Note: 2023 N=89,184; 2024 N=65,437. 受访者来自全球。来源：Stack Overflow 开发者调查（5/23 和 5/24-6/24）



# AI Developer Repositories – GitHub = ~175% Increase Over Sixteen Months

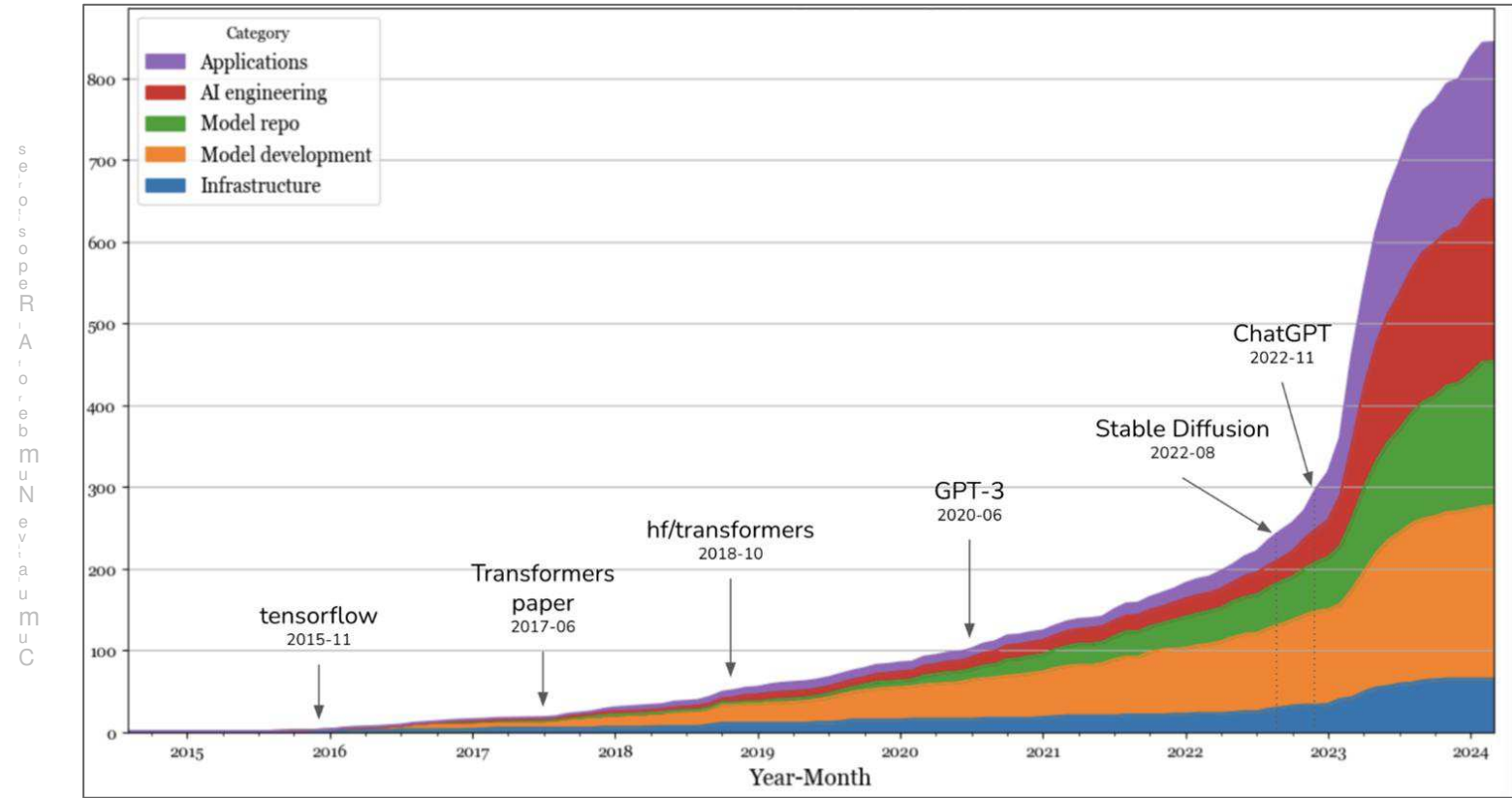
# AI 开发者仓库 – GitHub =~175 与十六个月前相比的增长率

Number of AI Developer Repositories\* on GitHub – 11/22-3/24, per Chip Hyuen



*\*A repository is an online storage space where developers share and manage code, models, data, and documentation related to artificial intelligence projects. These enable collaboration, reuse, and distribution of AI tools and assets. Analysis shown includes GitHub repositories with 500+ stars. Infrastructure = tools for model serving, compute management, vector search & databases. Model development = frameworks for modeling & training, inference optimization, dataset engineering, & model evaluation. Application development = custom AI-powered applications (varied use cases). Source: Chip Hyuen via GitHub (3/24)*

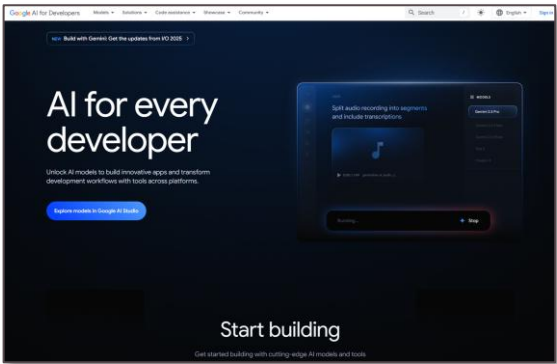
Number of AI Developer Repositories\* on GitHub – 11/22-3/24, per Chip Hyuen



*\* 仓库是一个在线存储空间，开发者可以在其中共享和管理与人工智能项目相关的代码、模型、数据和文档。这些仓库能够实现 AI 工具和资产的协作、重用和分发。显示的分析包括具有 500+ 星的 GitHub 仓库。用于模型服务、计算管理、向量搜索和数据库的基础设施 = 工具。用于建模和训练、推理优化、数据集工程和模型评估的模型开发 = 框架。应用程序开发 = 自定义的 AI 驱动应用程序（各种用例）。来源：Chip Hyuen，通过 GitHub （3/24）*

# AI Developer Ecosystem – Google = +50x Monthly Tokens Processed Y/Y

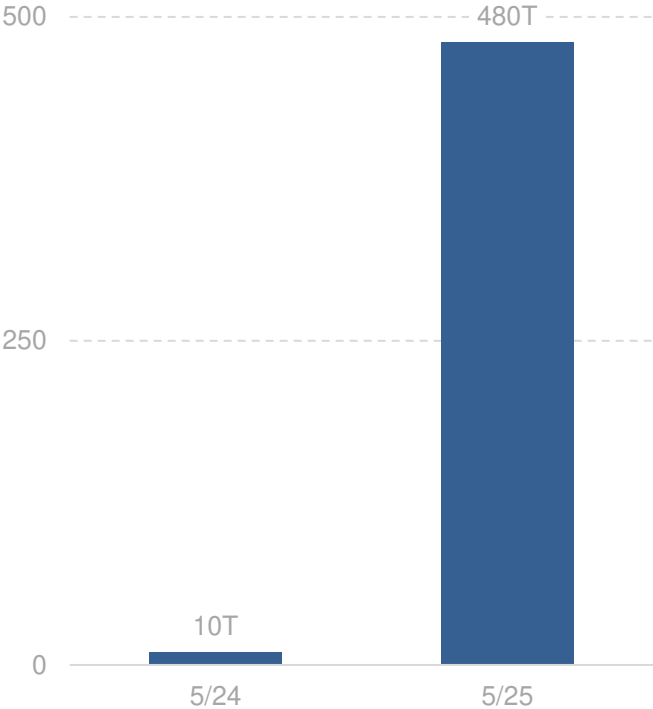
Google Monthly Tokens Processed (T) – 5/24-5/25, per Google



*This time last year, we were processing 9.7 trillion tokens a month across our products and APIs. Now, we’re processing over 480 trillion – that’s 50 times more.*

- Google I/O 2025 Press Release, 5/25

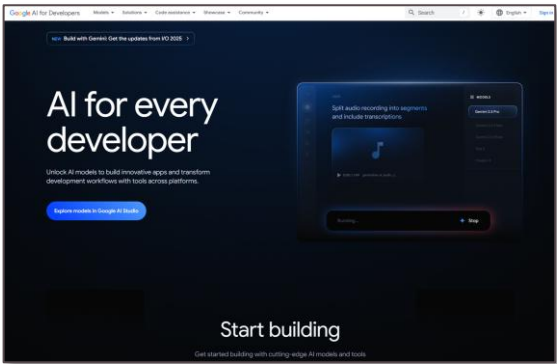
Monthly Tokens Processed, Trillions



Note: Token usage shown across Google products & APIs. Per Google in 5/25, ‘This time last year, we were processing 9.7 trillion tokens a month across our products and APIs. Now, we’re processing over 480 trillion — that’s 50 times more...Over 7 million developers are building with Gemini, five times more than this time last year.’ Source: Google, ‘Google I/O 2025: From research to reality’ (5/25)

# AI Developer Ecosystem – Google = +50x Monthly Tokens Processed Y/Y

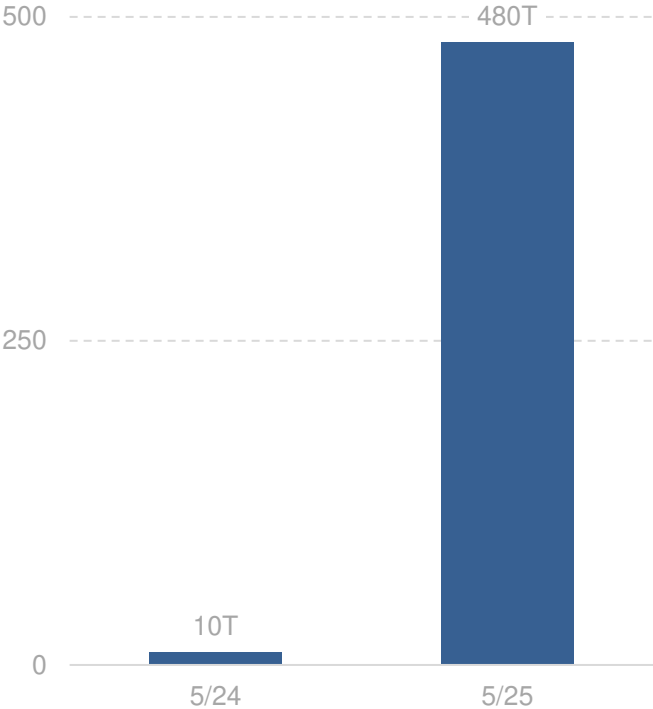
Google 每月处理的 Token (T) – 5/24-5/25, 来源：Google



去年这个时候，我们每月在我们的产品和 API 中处理 9.7 万亿个 token。现在，我们正在处理超过 480 万亿个 token —— 这是 50 倍以上。

- Google I/O 2025 新闻稿，5/25

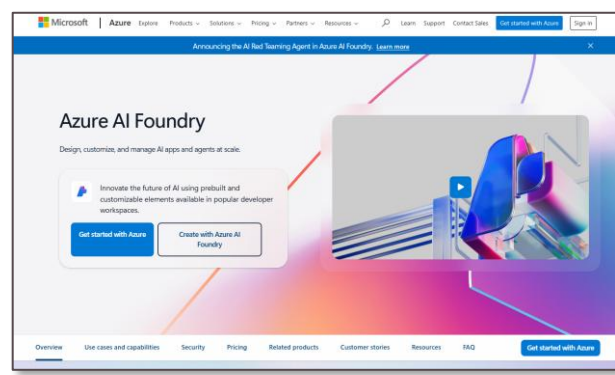
snorTdesecorPsnegotyhtnom



注意：显示的 Token 使用量涵盖 Google 产品和 API。根据 Google 在 5/25 的说法，“去年这个时候，我们每月在我们的产品和 API 中处理 9.7 万亿个 token。现在，我们正在处理超过 480 万亿个 token —— 这是 50 倍以上 …… 超过 700 万开发者正在使用 Gemini 进行构建，是去年这个时候的五倍以上。”资料来源：Google，“Google I/O 2025：从研究到现实”（5/25）

# AI Developer Ecosystem – Microsoft Azure AI Foundry = +5x Quarterly Tokens Processed Y/Y

Microsoft Azure AI Foundry Quarterly Tokens Processed (T) – Q1:24-Q1:25, per Microsoft

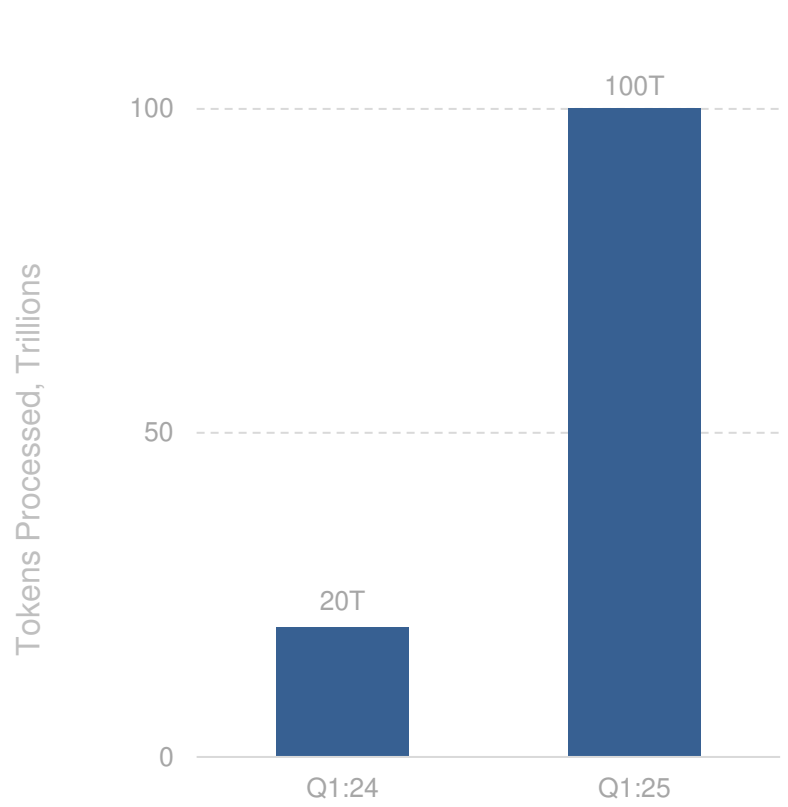


[Microsoft Azure AI] Foundry is the agent and AI app factory.

It is now used by developers at over 70,000 enterprises and digital natives – from Atomicwork, to Epic, Fujitsu, and Gainsight, to H&R Block and LG Electronics – to design, customize, and manage their AI apps and agents.

We processed over 100 trillion tokens this quarter, up 5x year-over-year – including a record 50 trillion tokens last month alone.

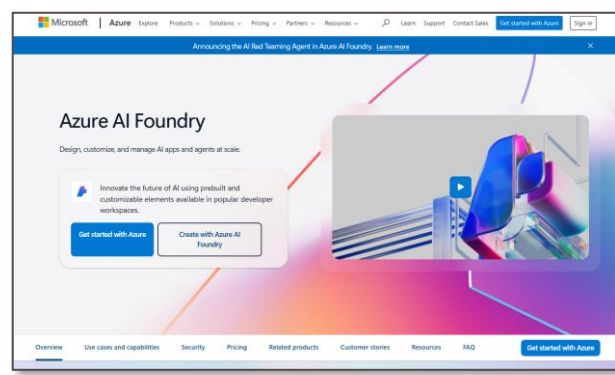
- Microsoft FQ3:25 Earnings Call, 4/25



Note: Source: Microsoft FQ3:25 earnings call (4/25)

# AI 开发者生态系统 – Microsoft Azure AI Foundry =+5x 季度代币处理量同比增长

Microsoft Azure AI Foundry Quarterly Tokens Processed (T) – Q1:24-Q1:25, per Microsoft

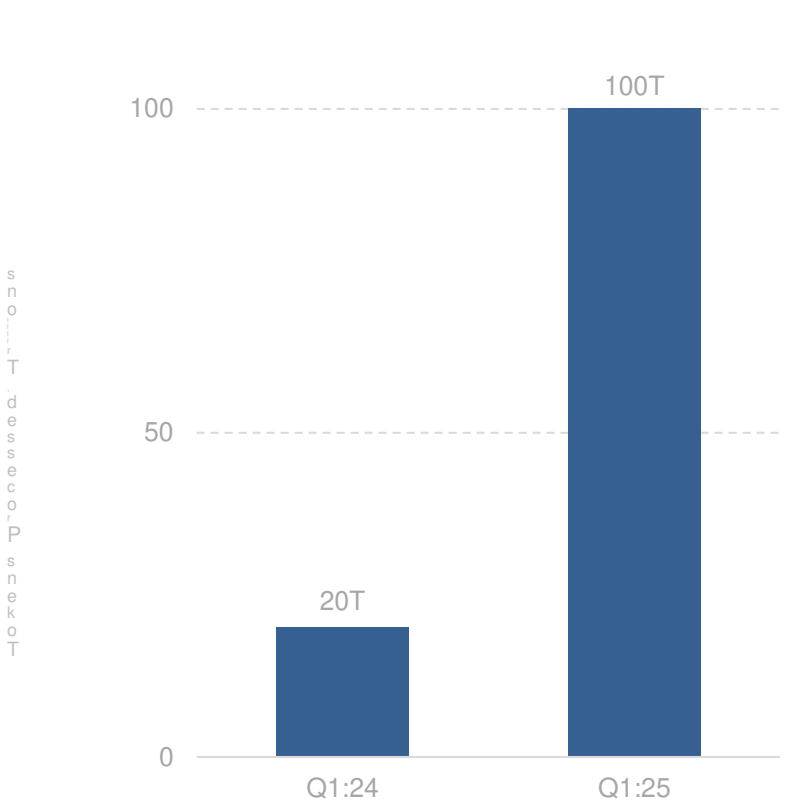


[Microsoft Azure AI] Foundry 是代理和 AI 应用工厂。

目前，超过 70,000 家企业和数字原生代的开发人员正在使用它 – 从 Atomicwork、Epic、Fujitsu 和 Gainsight，到 H&R Block 和 LG Electronics – 用于设计、定制和管理他们的 AI 应用和代理。

本季度我们处理了超过 100 万亿个 tokens，同比增长 5 倍 – 其中包括仅上个月就处理了创纪录的 50 万亿个 tokens。

- Microsoft FQ3:25 Earnings Call, 4/25


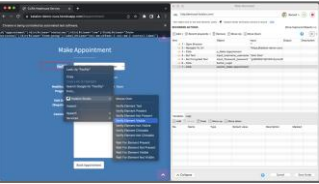
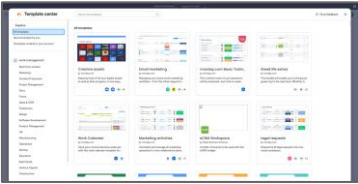


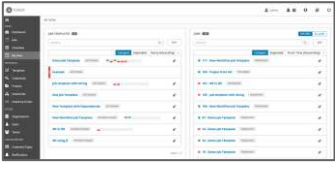
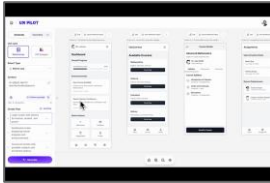
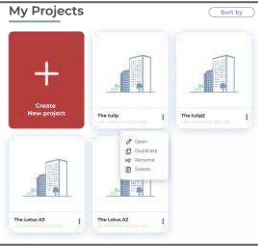


Note: Source: Microsoft FQ3:25 earnings call (4/25)

# AI Developer Use Cases = Broad & Varied


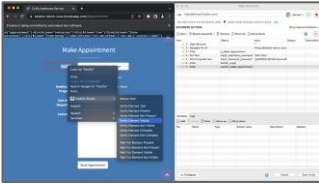

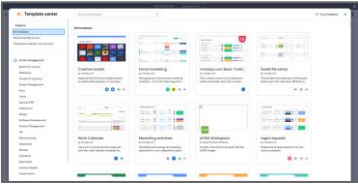


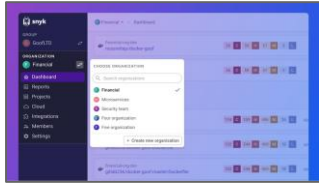
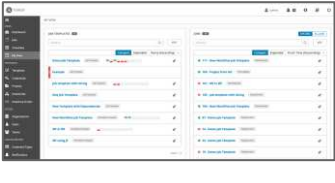
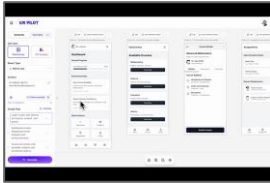
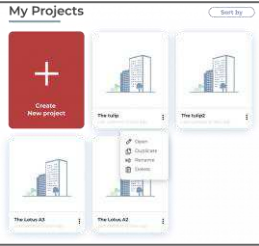
# AI 开发者用例 = 广泛且多样

AI Developer Use Cases – 2024, per IBM

<b>Code Generation</b> 	<b>Bug Detection &amp; Fixing</b> 	<b>Testing Automation</b> 	<b>Project / Workflow Management</b> 	<b>Documentation</b> 
<b>Refactoring &amp; Optimization</b> 	<b>Security Enhancement</b> 	<b>DevOps &amp; CI / CD Pipelines</b> 	<b>User Experience Design</b> 	<b>Architecture Design</b> 

Note: CI / CD pipelines are continuous integration / continuous deployment pipelines.  
Source: IBM, 'AI in Software Development' (2024); Anthropic; Katalon; AccelQ; Monday; Quill; Mintlify; Snyk; Ansible; UX Pilot; Ark Design AI

AI 开发者用例 – 2024, 根据 IBM

<b>Code Generation</b> 	<b>Bug Detection &amp; Fixing</b> 	<b>Testing Automation</b> 	<b>Project / Workflow Management</b> 	<b>Documentation</b> 
<b>Refactoring &amp; Optimization</b> 	<b>Security Enhancement</b> 	<b>DevOps &amp; CI / CD Pipelines</b> 	<b>User Experience Design</b> 	<b>Architecture Design</b> 

注意: CI / CD 管道是持续集成 / 持续部署管道。Source: IBM, 'AI in Software Development' (2024); Anthropic; Katalon; AccelQ; Monday; Quill; Mintlify; Snyk; Ansible; UX Pilot; Ark Design AI

*AI Model Compute Costs High / Rising*  
+  
*Inference Costs Per Token Falling*  
=  
*Performance Converging + Developer Usage Rising*  
  
*...(Likely) Long Way to Profitability*

*AI Model Compute Costs High / Rising*  
+  
每个 Token 的推理成本下降 =  
  
性能趋同 + 开发者使用量上升  
  
...（可能）实现盈利还有很长的路要走

- 1 Seem Like Change Happening Faster Than Ever?  
Yes, It Is
- 2 AI User + Usage + CapEx Growth =  
Unprecedented
- 3 AI Model Compute Costs High / Rising + Inference Costs Per Token Falling =  
Performance Converging + Developer Usage Rising
- 4 AI Usage + Cost + Loss Growth =  
Unprecedented
- 5 AI Monetization Threats =  
Rising Competition + Open-Source Momentum + China's Rise
- 6 AI & Physical World Ramps =  
Fast + Data-Driven
- 7 Global Internet User Ramps Powered by AI from Get-Go =  
Growth We Have Not Seen Likes of Before
- 8 AI & Work Evolution =  
Real + Rapid

- 1 变化似乎比以往任何时候都快？是的，的确如此
- 2 AI 用户 + 使用量 + 资本支出增长 = 前所未有
- 3 AI 模型计算成本高 / 不断上升 + 每次 Token 的推理成本下降 = 性能趋同 + 开发者使用量上升
- 4 AI 使用量 + 成本 + 亏损增长 = 前所未有
- 5 AI 货币化威胁 = 日益激烈的竞争 + 开源势头 + 中国的崛起
- 6 AI 与物理世界加速发展 = 快速 + 数据驱动
- 7 Global Internet User Ramps Powered by AI from Get-Go = 我们从未见过的增长速度
- 8 AI 与工作变革 = 真实的 + 迅速的

*It's different this time, we'll make it up on volume, and we'll figure out how to monetize our users in the future* are typically three of the biggest danger statements in business.

That said, in technology investing every once in awhile they can be gold – Amazon, Alphabet (Google), Meta (Facebook), Tesla, Tencent, Alibaba, Palantir...

In AI, it may indeed be different this time, and the leader(s) will make it up on volume and be able to monetize users in the future. Though now, ‘different this time’ also means that competition is unprecedented...

We have never seen so many founder-driven / assisted (ex. Apple) companies\* with market capitalizations in excess of \$1 trillion – most with gross margins of +50% plus free cash flow – attacking the same opportunity at the same time in a relatively transparent world, adding in high stakes competition between global powers – China and the United States.

Ernest Hemingway’s phrase *gradually, then suddenly* from ‘The Sun Also Rises’ applies to technology tipping points.

The tipping point for personal computers was the introduction of Apple’s Macintosh (1984) and Microsoft’s Windows 3.0 (1990). With the Internet it was Netscape’s IPO (1995). With the Mobile Internet it was Apple’s iPhone App Store launch (2008). With Cloud Computing it was the launch of AWS (Amazon Web Services) foundational products (2006-2009). With AI it was the launch of NVIDIA’s A100 GPU chip (2020) and OpenAI’s public version of ChatGPT (2022). In effect, the global competition for AI kicked in with the launch of China’s DeepSeek (1/25) and Jack Ma’s attendance at Chinese President Xi Jinping’s symposium of Chinese business leaders (2/25).

The money to fund AI’s growth (and losses) comes from big companies with big free cash flow and big balance sheets, in addition to wealthy and ambitious capital providers from around the world.

No doubt, this dynamic combination of competition / capital / entrepreneurship will rapidly advance AI, a riddle is determining which business models will be the last ones standing.

*\*Companies include NVIDIA, Microsoft, Amazon, Alphabet (Google), Meta (Facebook) & Tesla*

这次有所不同，我们将通过数量来弥补，而且我们将来会弄清楚如何将我们的用户货币化，这通常是商业中最大的三个危险声明。

也就是说，在技术投资中，偶尔它们也能成为金矿 —— 亚马逊、Alphabet (Google)、Meta (Facebook)、Tesla、Tencent、Alibaba、Palantir……

在人工智能领域，这次可能确实有所不同，领导者将通过数量来弥补，并且能够在未来将用户货币化。虽然现在，“这次有所不同”也意味着竞争是前所未有的……

我们从未见过如此多创始人驱动 / 辅助（例如 Apple）的公司\*，其市值超过 1 万亿美元 – 大多数公司的毛利率为 +50% 加上自由现金流 – 在相对透明的世界中，同时攻击相同的机会，再加上全球大国之间的高风险竞争 – 中国和美国。

欧内斯特 · 海明威在《太阳照常升起》中的短语逐渐地，然后突然地适用于技术引爆点。

个人电脑的引爆点是苹果公司推出 Macintosh (1984) 和微软公司推出 Windows 3.0 (1990)。互联网的引爆点是 Netscape 的首次公开募股 (1995)。移动互联网的引爆点是苹果公司 iPhone 应用商店的推出 (2008)。云计算的引爆点是 AWS (Amazon Web Services) 基础产品的推出 (2006-2009)。人工智能的引爆点是 NVIDIA 的 A100 GPU 芯片 (2020) 和 OpenAI 的 ChatGPT 公开版本 (2022) 的推出。实际上，随着中国 DeepSeek (1/25) 的推出以及马云出席中国国家主席习近平的中国商界领袖座谈会 (2/25)，全球人工智能竞争正式拉开帷幕。

为人工智能的增长（和亏损）提供资金的是拥有大量自由现金流和庞大资产负债表的大公司，以及来自世界各地富有且雄心勃勃的资本提供者。

毫无疑问，这种竞争 / 资本 / 企业家精神的动态结合将迅速推动人工智能的发展，但一个难题是确定哪些商业模式将最终胜出。

\* 公司包括 NVIDIA、Microsoft、Amazon、Alphabet (Google)、Meta (Facebook) 和 Tesla



# Technology Disruption Pattern Recognition = Hundreds of Years of Consistent Signals

# 技术颠覆模式识别 = 数百年来一致的信号

Technology disruption has a long-repeating rhythm: early euphoria, break-neck capital formation, bruising competition, and – eventually – clear-cut winners and losers.

Alasdair Nairn’s ‘Engines That Move Markets’ ([link here](#)) distills two centuries of such cycles, and his observations are prescient for today’s AI boom. Highlights of his observations follow...

*There were several years of strong share-price growth when the railways were supplanting canals. The bubble of the 1840s deflated under the weight of overheated expectations and changing economic conditions...*

*...Any technological advance which requires huge capital expenditure always runs a real risk of disappointing returns in the early years, even if it is ultimately successful...*

*...Any technology that necessitates heavy capital expenditure and requires returns to be earned over an extended period is always going to be a high-risk undertaking – unless, that is, there is some form of protection against competition...*

*...The winners of these competitive struggles are not always those who have the best technology, but those who can most clearly see the way that an industry or market is likely to develop...*

*...One of the clearest lessons of corporate and investment history is that without some barrier to entry, first-mover advantage can be swiftly lost...*

*...A theme that recurs throughout this research is that while identifying the winners from any new technology is often perilous and difficult, it is almost invariably simpler to identify who the ‘losers’ are going to be.*

技术颠覆具有长期重复的节奏：早期欣快感、惊人的资本形成、激烈的竞争，以及 – 最终 – 明确的赢家和输家。

阿拉斯代尔 · 奈恩的《推动市场的引擎》（[link here](#)）提炼了两个世纪以来的此类周期，他的观察对于今天的 AI 繁荣具有先见之明。以下是他观察的重点 ……

当铁路取代运河时，曾有几年的股价强劲增长。19 世纪 40 年代的泡沫在过热的预期和不断变化的经济状况的重压下破灭 ……

… 任何需要巨额资本支出的技术进步，即使最终成功，在早期也总是存在令人失望的回报的真正风险 ……

… 任何需要大量资本支出并需要在较长时间内获得回报的技术，始终是一项高风险的 undertaking – 除非有某种形式的保护来对抗竞争 ……

… 这些竞争的胜利者并不总是那些拥有最佳技术的人，而是那些能够最清楚地看到一个行业或市场可能发展方向的人 ……

… 公司和投资历史中最清晰的教训之一是，如果没有某种进入壁垒，先发优势可能会迅速丧失 ……

… 贯穿这项研究的一个主题是，虽然从任何新技术中识别出赢家通常是危险和困难的，但几乎总是更容易识别出谁将是“失败者”。



*AI-Related Monetization =  
Very Robust Ramps*

与人工智能相关的货币化 =  
非常强大的斜坡

To understand the evolution of AI hardware strategy, we'll look at how control over chip design is shifting from traditional vendors to the platforms that rely on them. For years, NVIDIA has been at the center of the AI hardware stack. Its GPUs (graphics processing units) became the default engine for training and inference, prized for their ability to handle highly parallel computations at scale. Its proprietary technology – and unparalleled scale – has led to industry leadership.

This reliance – combined with outsized sudden demand – has created constraints. Despite NVIDIA's rapid – and impressive – scale-up, demand for NVIDIA GPUs has outpaced supply amid industry fervor for accelerated computing. Hyperscalers and cloud providers are moving to improve their supply chains to manage long lead times.

That shift is accelerating the rise of custom silicon – especially ASICs, or application-specific integrated circuits. Unlike GPUs, which are designed to support a wide range of workloads, ASICs are purpose-built to handle specific computational tasks with maximum efficiency. In AI, that means optimized silicon for matrix multiplication, token generation, and inference acceleration.

Google's TPU (Tensor Processing Unit) and Amazon's Trainium chips are now core components of their AI stacks. Amazon claims its Trainium2 chips offer 30-40% better price-performance than standard GPU instances, unlocking more affordable inference at scale. These aren't side projects – they're foundational bets on performance, economics, and architectural control...

为了理解人工智能硬件策略的演变，我们将了解芯片设计的控制权是如何从传统供应商转移到依赖它们的平台的。多年来，NVIDIA 一直是人工智能硬件堆栈的中心。它的 GPU（图形处理单元）已成为训练和推理的默认引擎，因其能够大规模处理高度并行计算而备受推崇。它的专有技术 – 和无与伦比的规模 – 使其在行业中处于领先地位。

这种依赖性 – 加上超乎寻常的突发需求 – 造成了限制。尽管 NVIDIA 迅速且令人印象深刻地 – 扩大了规模，但在行业对加速计算的热情中，对 NVIDIA GPU 的需求超过了供应。超大规模企业和云提供商正在努力改善其供应链，以管理较长的交货时间。

这种转变正在加速定制芯片的崛起 – 尤其是 ASIC，即专用集成电路。与旨在支持各种工作负载的 GPU 不同，ASIC 专门用于以最高的效率处理特定的计算任务。在人工智能领域，这意味着针对矩阵乘法、令牌生成和推理加速进行优化的芯片。

谷歌的 TPU（张量处理器）和亚马逊的 Trainium 芯片现在是其 AI 堆栈的核心组件。亚马逊声称其 Trainium2 芯片提供比标准 GPU 实例高 30-40% 的性价比，从而以更经济实惠的价格大规模解锁推理。这些不是辅助项目 – 它们是对性能、经济性和架构控制的基础性押注 ……

...Custom chips also reflect a broader effort to manage the economics of AI infrastructure. As Amazon CEO Andy Jassy noted in early 2025, *AI does not have to be as expensive as it is today, and it won't be in the future*. Custom silicon is one lever to control these expenses.

At the same time, a new ecosystem of infrastructure specialists is emerging to meet this demand.

CoreWeave has become one of the fastest-scaling cloud GPU providers, repurposing gaming and Crypto hardware supply chains to serve enterprise AI customers.

Oracle, long seen as a legacy IT vendor, has repositioned itself as a GPU-rich cloud platform with AI-specific offerings. Astera Labs, a lesser-known but critical player, builds high-speed interconnects that move data between GPUs and memory systems with minimal latency – an increasingly important performance constraint.

These firms aren't building foundation models, but they're building what foundation models depend on.

As compute demand compounds, they're becoming essential infrastructure in a market where speed, availability, and efficiency are important differentiators.

… 定制芯片也反映出管理人工智能基础设施经济效益的更广泛努力。正如亚马逊首席执行官 Andy Jassy 在 2025 年初指出的那样，人工智能不必像今天这样昂贵，而且未来也不会如此。定制芯片是控制这些费用的一个手段。

与此同时，一个新的基础设施专家生态系统正在涌现，以满足这一需求。

CoreWeave 已成为发展最快的云 GPU 提供商之一，它将游戏和加密硬件供应链重新用于服务企业 AI 客户。

长期以来被视为传统 IT 供应商的 Oracle，已将自己重新定位为一个拥有丰富 GPU 的云平台，并提供特定于 AI 的产品。Astera Labs 是一家鲜为人知但至关重要的参与者，它构建高速互连，以最小的延迟在 GPU 和内存系统之间移动数据 – 这是一个日益重要的性能约束。

These firms aren't building foundation models, but they're building what foundation models depend on.

随着计算需求的增加，它们正在成为一个市场中的重要基础设施，在这个市场中，速度、可用性和效率是非常重要的区别因素。

AI Monetization =

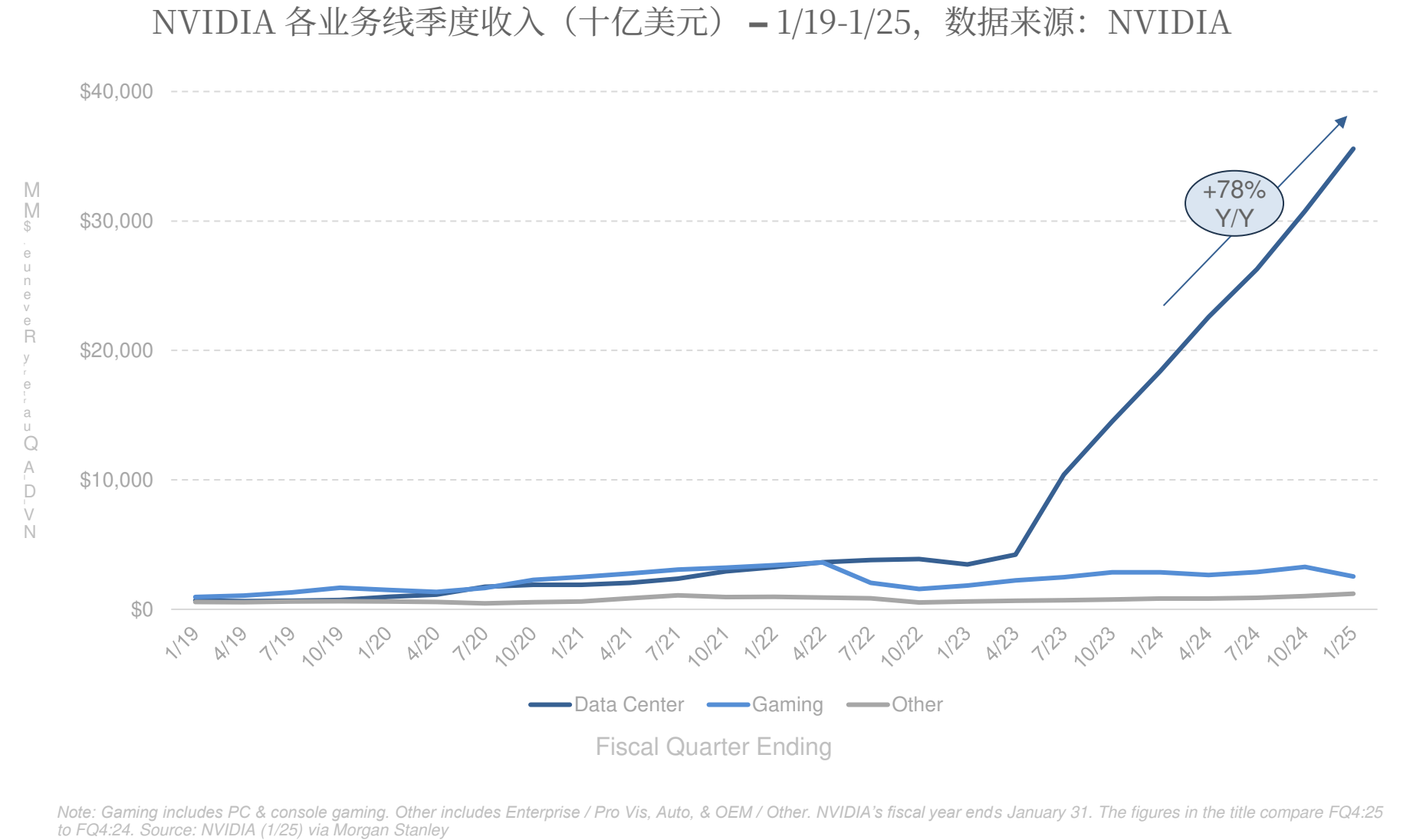
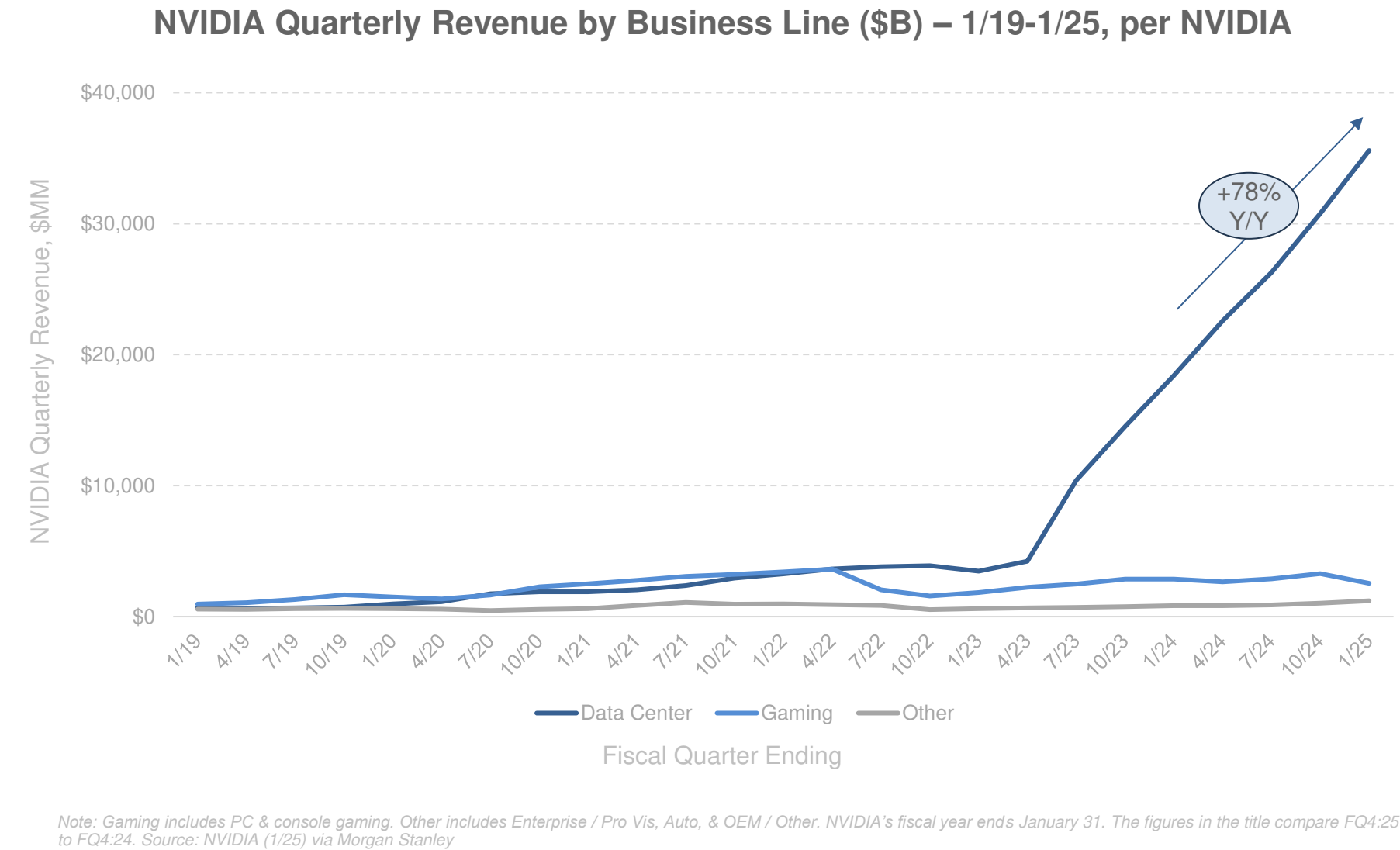
Chips

AI 货币化 =

芯片

AI Monetization...Chips =  
NVIDIA Quarterly Revenue +78% to \$39B Y/Y...

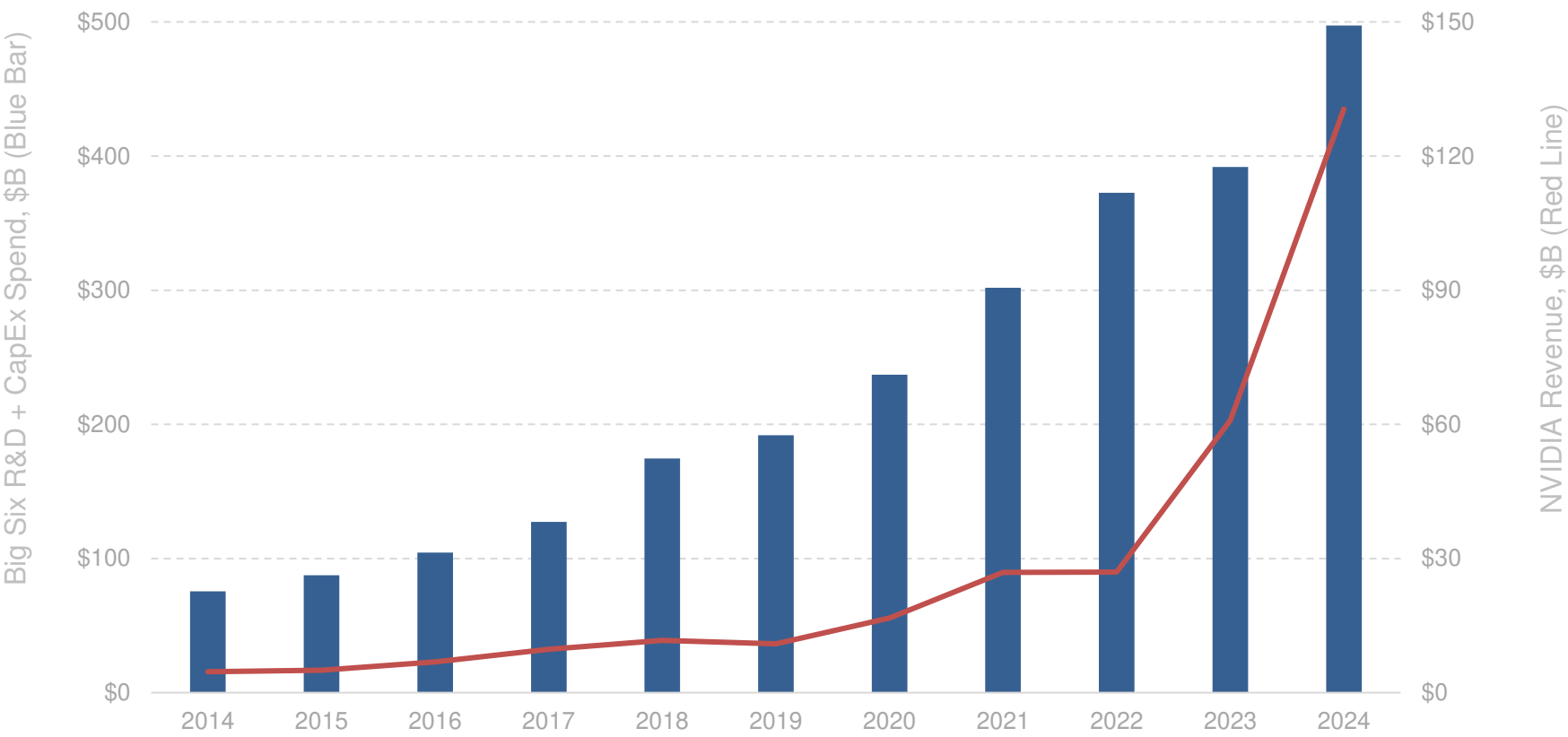
AI 货币化 … 芯片 =NVIDIA 季度收入 +78% 至 390 亿  
美元（同比增长） …



...AI Monetization...Chips =  
NVIDIA Revenue +28x Over Ten Years...Big Six CapEx + R&D +6x

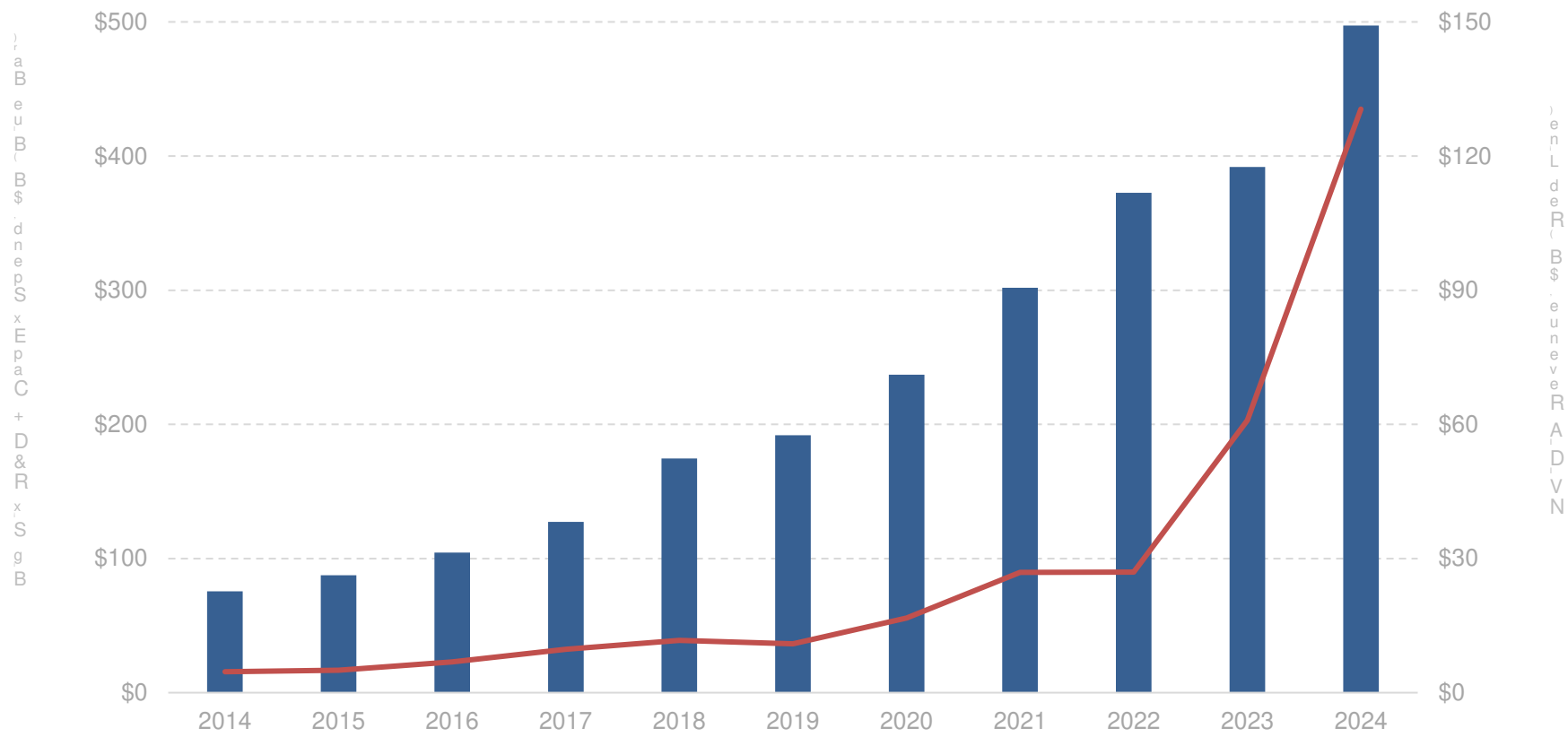
…AI 货币化 … 芯片 =NVIDIA 收入 +28 十年来的倍数 … 六巨头 资本支出 +  
研发 +6 倍

Big Six\* USA Public Technology Company R&D + CapEx Spend (\$B)  
vs. NVIDIA Revenue (\$B) – 2014-2024, per Capital IQ



\*Note: Big Six USA technology companies include Apple, Nvidia, Microsoft, Alphabet / Google, Amazon, & Meta Platforms / Facebook. Includes CapEx for Amazon AWS + Retail as R&D expense is not regularly separated for those two business divisions. Source: Companies' investor reports, Capital IQ (4/25)

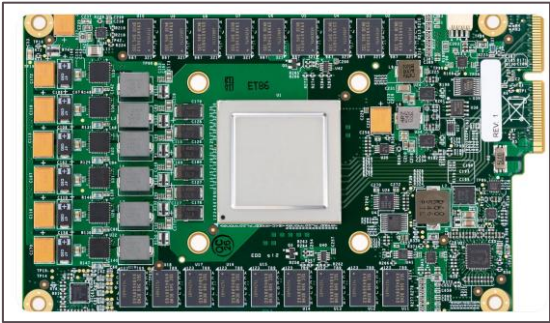
美国六大 \* 公开科技公司研发 + 资本支出（10 亿美元）与 NVIDIA 收入（10 亿美元）对比 – 2014-2024，数据来源：Capital IQ



\* 注：美国六大科技公司包括 Apple、Nvidia、Microsoft、Alphabet / Google、Amazon 和 Meta Platforms / Facebook。包括 Amazon AWS + 零售的资本支出，因为研发费用并未针对这两个业务部门定期单独列出。来源：公司投资者报告，Capital IQ (4/25)

# AI Monetization...Chips = Google TPU Sales\* +116% to \$8.9B Y/Y, per Morgan Stanley

Google TPU (Tensor Processing Unit) Estimated Sales – 2021-2024, per Morgan Stanley

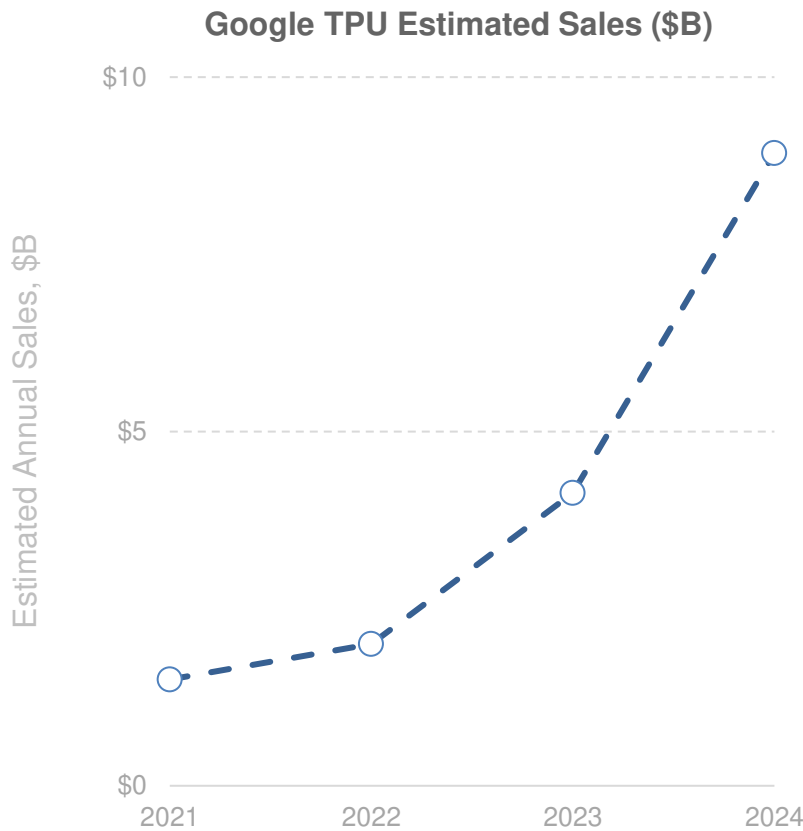


TPUs were purpose-built specifically for AI. TPUs are an application-specific integrated circuit (ASIC), a chip designed for a single, specific purpose: running the unique matrix and vector-based mathematics that's needed for building and running AI models.

Our first such chip, TPU v1, was deployed internally in 2015 and was instantly a hit across different parts of Google...

...‘We thought we'd maybe build under 10,000 of them,’ said Andy Swing, principal engineer on our machine learning hardware systems. ‘We ended up building over 100,000.’

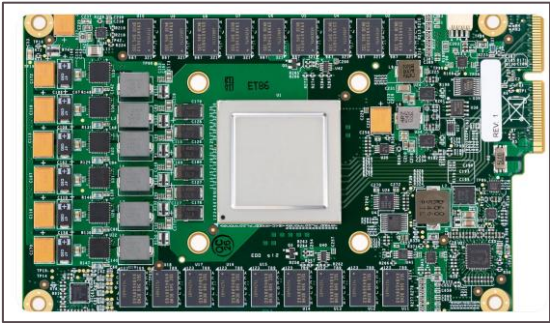
- Google Press Release, 7/24



\*Figures are estimates per Morgan Stanley research. Note: Relative to GPUs, ASICs are custom-designed for specific tasks (e.g., AI model training,) whereas GPUs are general-purpose. Source: Google, Morgan Stanley, ‘GenAI Monetization – Assessing The ROI Equation’ (2/25)

# AI Monetization...Chips =Google TPU Sales\* +116% 至 \$8.9B Y/Y, 据摩根士丹利称

Google TPU (Tensor Processing Unit) 预估销量 – 2021-2024, 据摩根士丹利

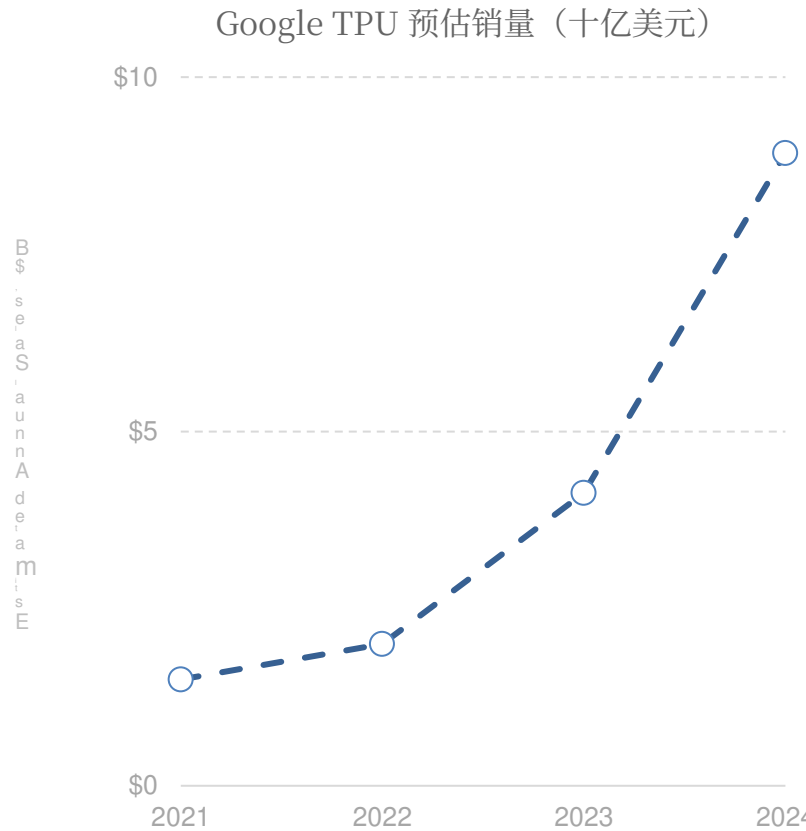


TPU 专门为 AI 构建。TPU 是一种专用集成电路 (ASIC)，这是一种为单一、特定目的而设计的芯片：运行独特的矩阵和矢量 - 基于构建和运行 AI 模型所需的数学。

我们的第一个此类芯片 TPU v1 于 2015 年在内部署并立即在 Google 的不同部门中大受欢迎 ……

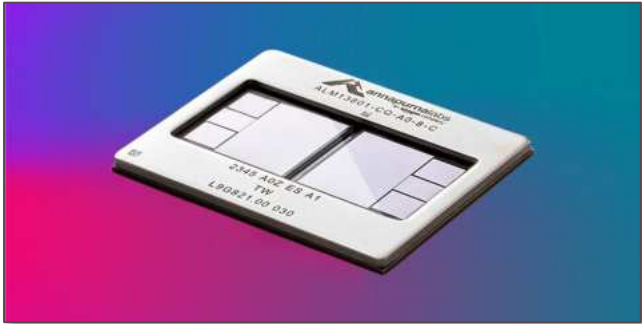
……“我们认为我们可能会建造不到 10,000 个，”我们的机器学习首席工程师 Andy Swing 说硬件系统。“我们最终建造了超过 100,000 个。”

- Google 新闻稿, 7/24



\*数字是根据摩根士丹利研究估算的。注意：相对于 GPU，ASIC 是为特定任务（例如，AI 模型训练）定制设计的，而 GPU 是通用 - 用途。来源：Google、摩根士丹利，《GenAI 货币化——评估 ROI 方程》（2/25）

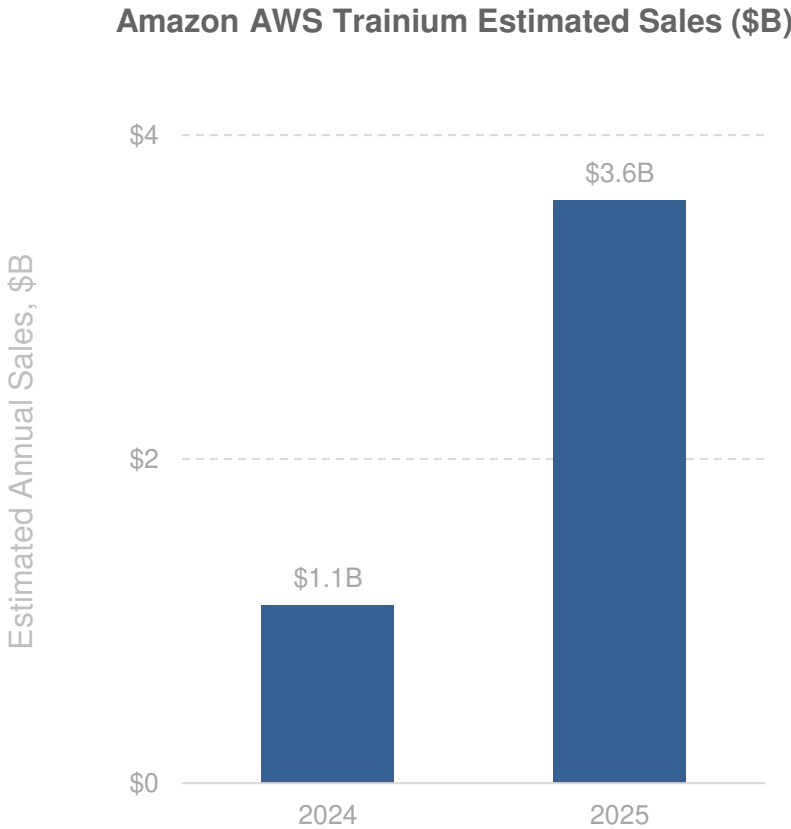
Amazon AWS Trainium Estimated Sales – 2024-2025, per Morgan Stanley



AWS Trainium chips are a family of AI chips purpose built by AWS for AI training and inference to deliver high performance while reducing costs...

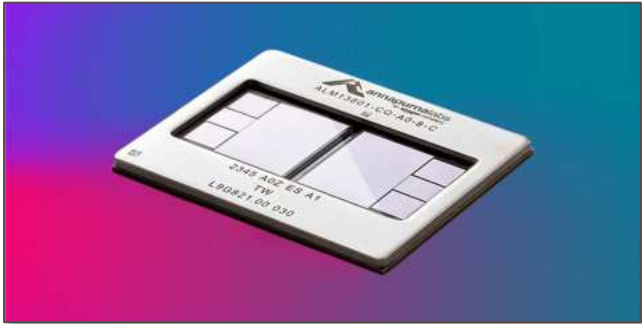
AWS Trainium2 chip delivers up to 4x the performance of first-generation Trainium...[and offers] 30-40% better price performance than the current generation of GPU-based EC2 P5e and P5en instances.

- Amazon AWS Trainium Overview, Accessed 5/25



Note: Relative to GPUs, ASICs are custom-designed for specific tasks (e.g., AI model training,) whereas GPUs are general-purpose. Figures are estimates per Morgan Stanley research. Source: Amazon AWS, Morgan Stanley, 'GenAI Monetization – Assessing The ROI Equation' (2/25)

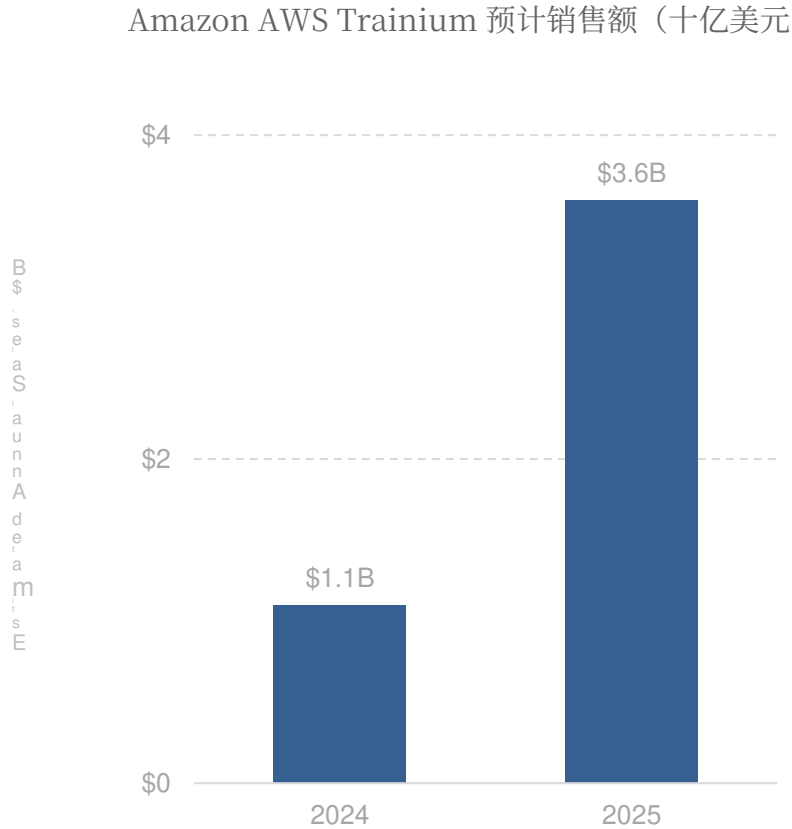
Amazon AWS Trainium Estimated Sales – 2024-2025, per Morgan Stanley



AWS Trainium 芯片是 AWS 专门为 AI 训练和推理构建的一系列 AI 芯片，旨在提供高性能，同时降低成本 ……

AWS Trainium2 芯片的性能是第一代 Trainium 的 4 倍 ……[并提供 ] 30 比当前一代基于 GPU 的 EC2 P5e 和 P5en 实例高 40% 的性价比。

- Amazon AWS Trainium 概述，访问时间 5/25



注意：相对于 GPU，ASIC 是为特定任务（例如，AI 模型训练）定制设计的，而 GPU 是通用型的。数字是摩根士丹利研究的估计值。来源：Amazon AWS、摩根士丹利，《GenAI Monetization – Assessing The ROI Equation》（2/25）



AI Monetization =  
Compute Services

AI 货币化 =  
Compute Services

# AI Monetization...Cloud Computing = CoreWeave Revenue +730% to \$1.9B Y/Y

CoreWeave Revenue – 2022-2024, per CoreWeave



We've delivered an outstanding start to 2025 on multiple fronts. Our strong first quarter financial performance caps a string of milestones including our IPO, our major strategic deal with OpenAI as well as other customer wins, our acquisition of Weights & Biases and many technical achievements...

...Demand for our platform is robust and accelerating as AI leaders seek the highly performant AI cloud infrastructure required for the most advanced applications. We are scaling as fast as possible to capture that demand. The future runs on CoreWeave.

- CoreWeave CEO Michael Intrator, 5/25

Source: CoreWeave (as of 5/25)

# AI Monetization...Cloud Computing = CoreWeave Revenue +730% to \$1.9B Y/Y

CoreWeave 营收 –2022-2024, 数据来源: CoreWeave



我们在多个方面为 2025 年取得了出色的开局。我们强劲的第一季度财务业绩为一系列里程碑画上了圆满的句号，包括我们的 IPO、与 OpenAI 达成的重大战略协议以及其他客户的成功、我们对 Weights & Biases 和许多技术成就的收购 ……

…… 对我们平台的需求强劲且正在加速增长，因为 AI 领导者正在寻求最先进的应用程序所需的高性能 AI 云基础设施。我们正在尽可能快地进行扩展以满足这一需求。未来在 CoreWeave 上运行。

- CoreWeave 首席执行官 Michael Intrator, 5/25

来源: CoreWeave (截至 5 月 25 日)

# AI Monetization...AI Infrastructure = Oracle Revenue +50x to \$948MM Over Two Years

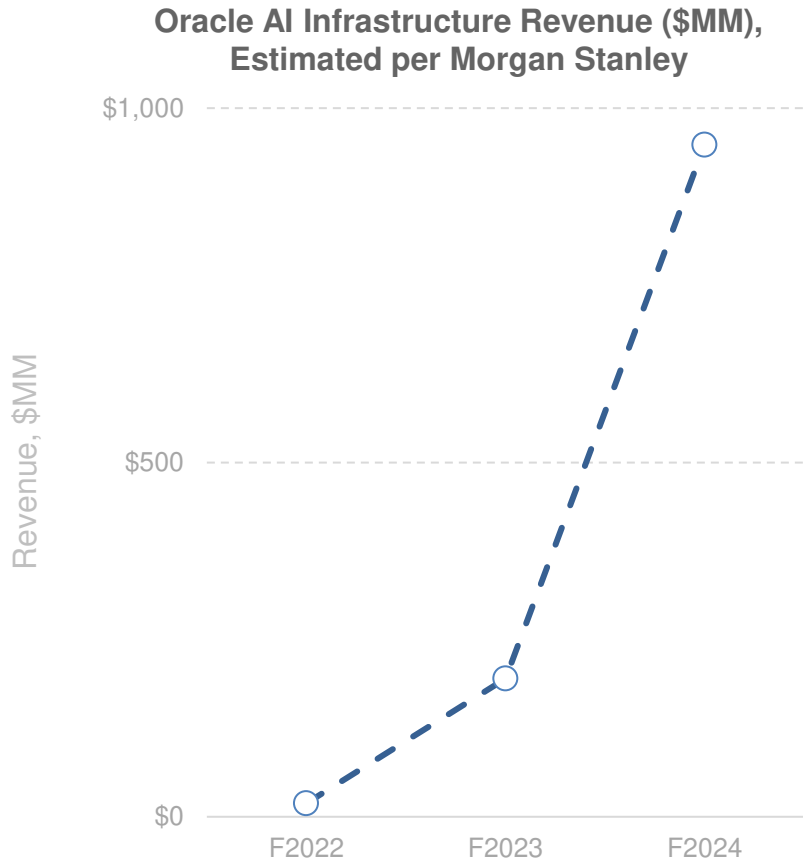
## Oracle AI Infrastructure Revenue – F2022-F2024, per Oracle & Morgan Stanley Estimates



There are many, many [AI infrastructure] customers who have come on and that haven't gotten capacity yet...

...We've got at least 40 new AI bookings that are over a billion (dollars) that haven't come online yet.

- Oracle CEO Safra Catz, 3/24



Source: Oracle, Morgan Stanley estimates, 'What's Ahead for the AI Infrastructure Cycle' (8/24)

# AI Monetization...AI Infrastructure = Oracle Revenue +50x to \$948MM Over Two Years

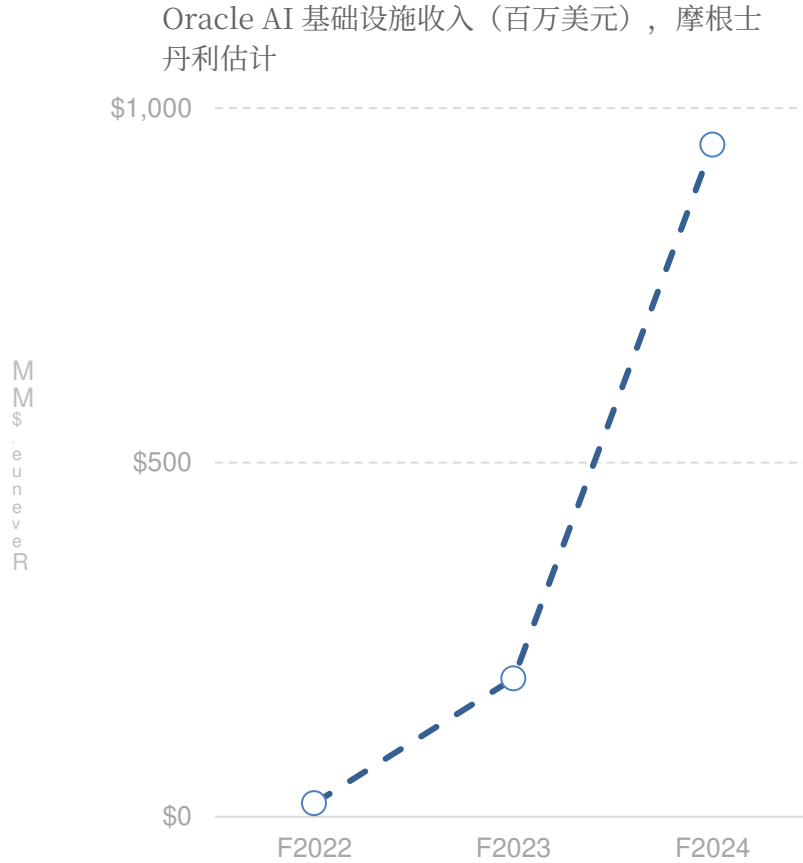
## Oracle AI Infrastructure Revenue – F2022-F2024, 根据 Oracle 和摩根士丹利估计



有非常多的 [AI infrastructure] 客户已经上线，但尚未获得容量 ...

…… 我们至少有 40 个新的 AI 预订，超过 10 亿（美元），但尚未上线。

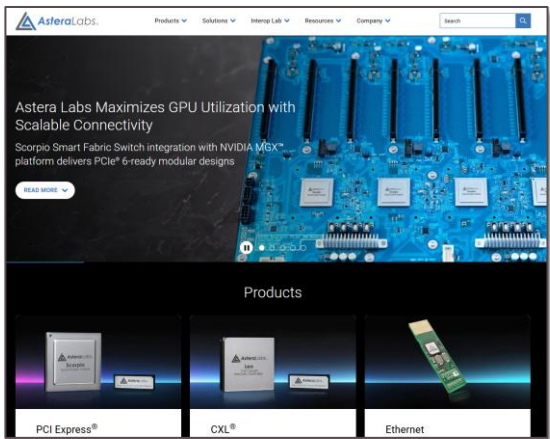
- Oracle CEO Safra Catz, 3/24



来源：Oracle，摩根士丹利估计，《人工智能基础设施周期的前景》（8/24）

# AI Monetization...Infrastructure Connectivity = Astera Labs Revenue +242% to \$396MM Y/Y

Astera Labs – 2022-2024, per Astera Labs

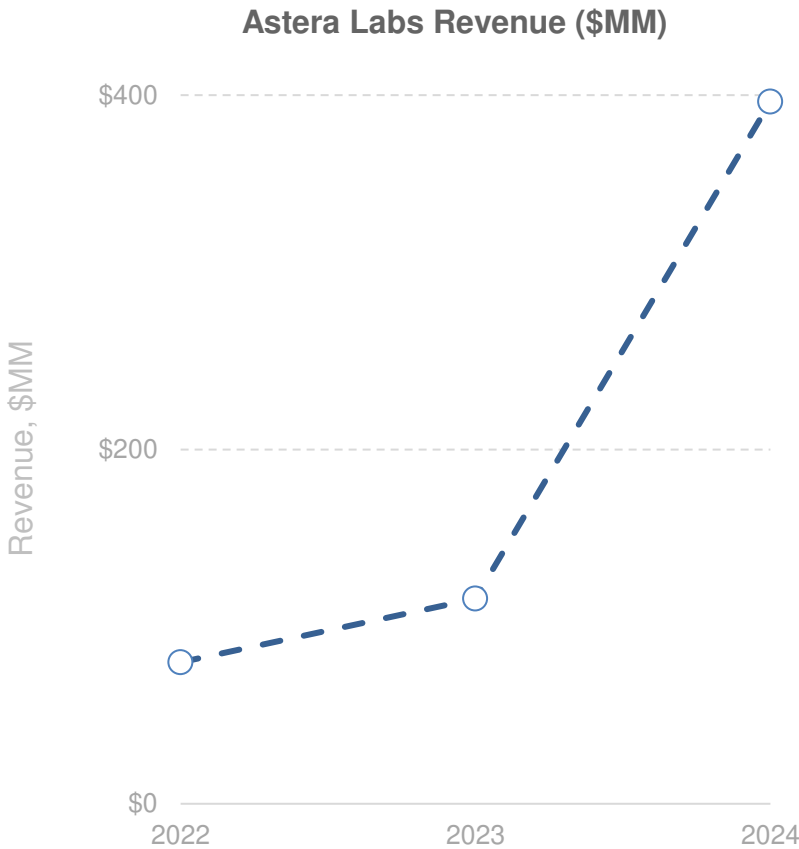


Astera Labs delivered strong Q4 results, with revenue growing 25% versus the previous quarter, and capped off a stellar 2024 with 242% revenue growth year-over-year...

...We expect 2025 to be a breakout year as we enter a new phase of growth driven by revenue from all four of our product families to support a diverse set of customers and platforms.

This includes our flagship Scorpio Fabric products for head-node PCIe connectivity and backend AI accelerator scale-up clustering.

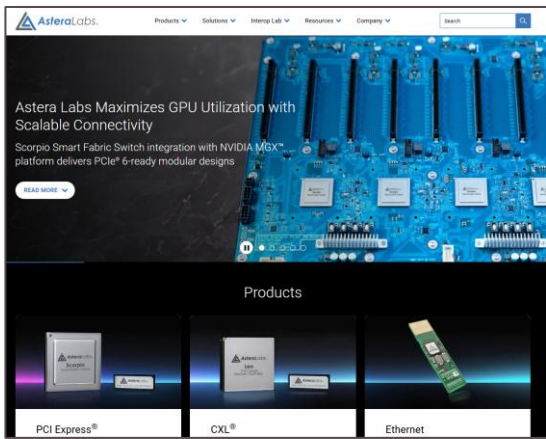
- Astera Labs CEO Jitendra Mohan, 2/25



Source: Astera Labs financial results (as of 4/25)

# AI 货币化 ... 基础设施连接 =Astera Labs 收入 +242% 至 \$3.96 亿美元（同比增长）

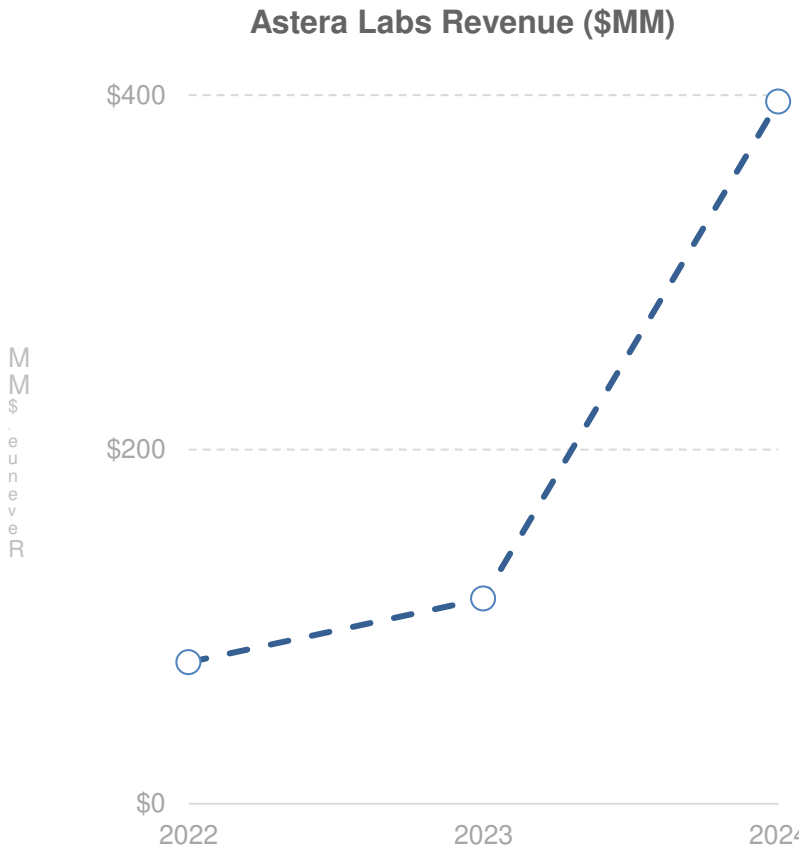
Astera Labs – 2022-2024, 数据来源：Astera Labs



Astera Labs 交付了强劲的第四季度业绩，收入环比增长 25%，并以 242% 的收入同比增长结束了 2024 年的辉煌业绩 ...

... 我们预计 2025 年将是突破性的一年，因为我们进入了一个新的增长阶段，该阶段由我们所有四个产品系列的收入驱动，以支持不同的客户和平台。这包括我们的旗舰 Scorpio Fabric 产品，用于头节点 PCIe 连接和后端 AI 加速器扩展集群。

- Astera Labs 首席执行官 Jitendra Mohan, 2 月 25 日



Source: Astera Labs financial results (as of 4/25)

# AI Monetization...Data Collection + Supercomputing = Tesla AI Training Capacity +8.5x

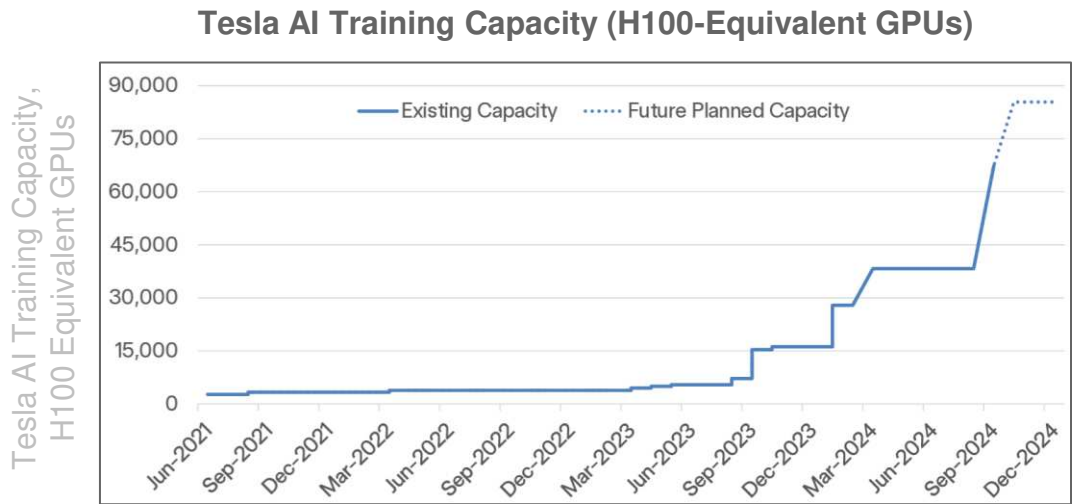
## Tesla Dojo Custom Supercomputer – 6/21-9/24, per Tesla



*We're continuing to simultaneously make significant purchases of GPUs and also putting a lot of effort into Dojo [custom supercomputer], which we believe has the potential for an order of magnitude improvement in the cost of training...*

*...Dojo also has the potential to become a sellable service that we would offer to other companies, in the same way that Amazon Web Services offers more web services, even though it started out as a bookstore. So, I really think that the Dojo potential is very significant.*

**- Tesla Co-Founder & CEO Elon Musk, 4/23**



Tesla AI Training Capacity,  
H100 Equivalent GPUs

Note: Listing capacity in 'H100-equivalent GPUs' means Tesla converts the aggregate AI-training throughput of Dojo and its other accelerators into the number of NVIDIA Hopper H100 data-center GPUs that would deliver the same FP8/FP16 FLOPS, giving a single, industry-standard yard-stick for compute scale.  
Source: Tesla Q1:23 earnings call, Tesla Q3:24 investor presentation, Data Center Dynamics, Wikimedia Commons

# AI 货币化 ... 数据收集 + 超级计算 = Tesla AI 训练容量 +8.5x

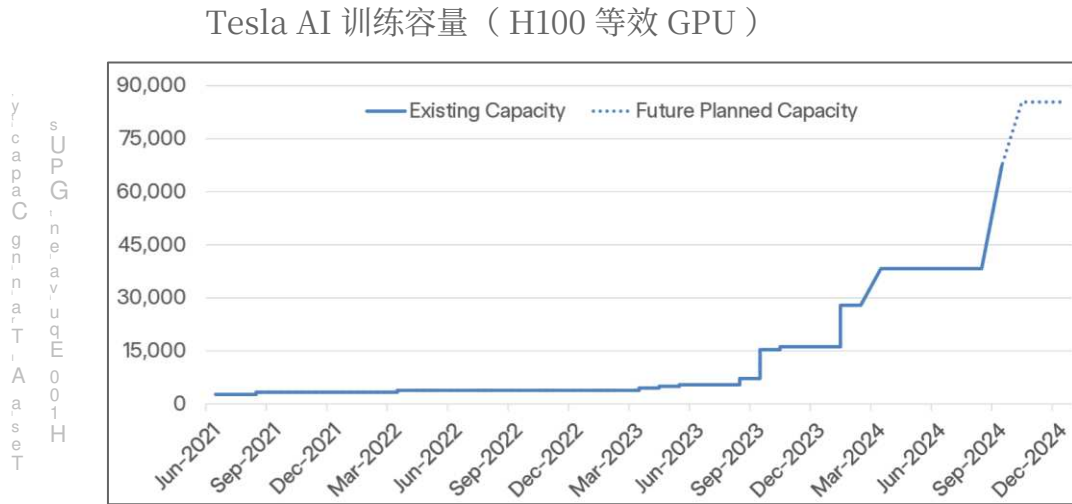
## Tesla Dojo 定制超级计算机 – 6/21-9/24, 按 Tesla



我们继续同时进行大量购买 GPU，并且还在 Dojo [定制超级计算机] 上投入大量精力，我们认为 Dojo 有可能在训练成本方面实现一个数量级的提升 .....

...Dojo 也有潜力成为一项可销售的服务，我们可以将其提供给其他公司，就像 Amazon Web Services 提供更多 Web 服务一样，尽管它最初是一家书店。所以，我真的认为 Dojo 的潜力非常巨大。

- Tesla 联合创始人兼 CEO Elon Musk, 4/23



Tesla AI Training Capacity,  
H100 Equivalent GPUs

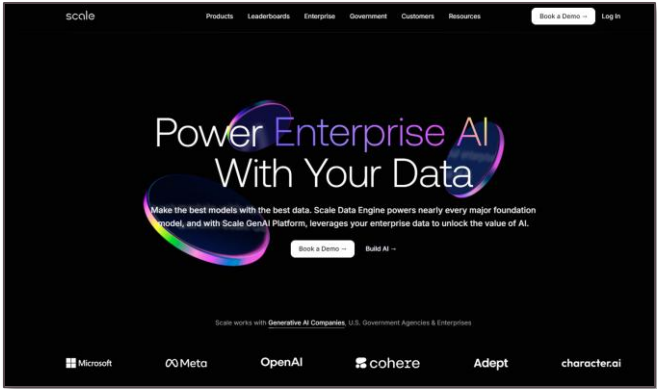
注意：以“H100-等效 GPU”列出容量意味着 Tesla 将 Dojo 及其它加速器的聚合 AI-训练吞吐量转换为可提供相同 FP8/FP16 FLOPS 的 NVIDIA Hopper H100 数据中心 GPU 的数量，从而为计算规模提供统一的行业标准衡量标准。来源：Tesla Q1:23 财报电话会议、Tesla Q3:24 投资者演示文稿、数据中心动态、维基共享资源

AI Monetization =  
Data Layer

AI 货币化 =  
数据层

# AI Monetization...Data Labeling & Evaluation = Scale AI Revenue +160% to \$870MM Y/Y

Scale AI Revenue – 2023-2024, per Scale AI



*Data abundance is not the default; it's a choice. It requires bringing together the best minds in engineering, operations, and AI. Our vision is one of data abundance, where we have the means of production to continue scaling frontier LLMs many more orders of magnitude. We should not be data-constrained in getting to GPT-10.*

**- Scale AI Co-Founder & CEO Alexandr Wang, 5/24**

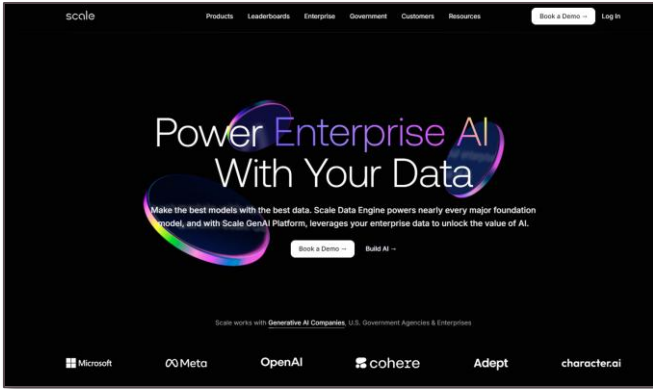
*We saw 160% revenue growth in 2024 from the previous year, and we secured more than \$1.5 billion in new business.*

**- Scale AI Head of Corporate and Product Comms Joe Osborne, 4/25**

*Note: 2023 figures are estimates based on Joe Osborne (Head of Corporate and Product Comms at Scale AI,) who indicated, 'We saw 160% revenue growth in 2024 from the previous year, and we secured more than \$1.5 billion in new business.' Source: Scale AI, The Information (4/25) ([link](#))*

# AI 货币化 … 数据标注与评估 =Scale AI Revenue +160% to \$870MM Y/Y

Scale AI Revenue – 2023-2024, 数据来源：Scale AI



数据丰富不是默认状态；而是一种选择。它需要汇集工程、运营和 AI 领域最优秀的人才。我们的愿景是数据丰富，我们拥有生产资料，可以继续将前沿 LLM 扩展多个数量级。在获得 GPT-10 的过程中，我们不应该受到数据限制。

- Scale AI 联合创始人兼首席执行官 Alexandr Wang, 5/24

2024 年，我们的收入同比增长了 160%，并获得了超过 15 亿美元的新业务。

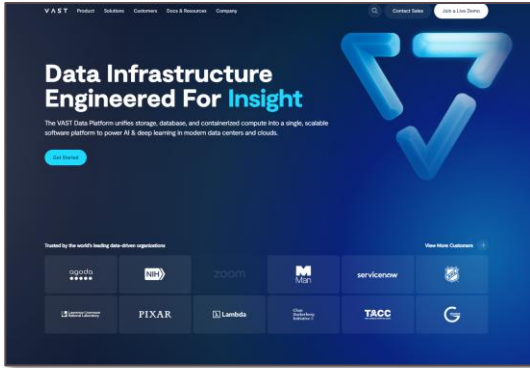
- Scale AI 公司与产品传播负责人 Joe Osborne, 4/25

注意：2023 年的数据是基于 Joe Osborne（Scale AI 的公司和产品传播负责人）的估计，他表示，“我们在 2024 年的收入增长率比上一年增长了 160%，并且我们获得了超过 15 亿美元的新业务。”来源：Scale AI，《The Information》（4/25）（链接）



# AI Monetization...Data Storage / Management / Processing = VAST Data Lifetime Sales From 0 to \$2B in Just Over Six Years

VAST Data – 1/19-5/25, per VAST Data



*Everything is accelerating. The rate of AI progress is constantly increasing as model builders build on each other's discoveries and push the boundaries ever farther. While we've been talking about thinking machines since early 2022, the advent of reasoning models in the last 12 months means that the era of thinking machines is actually now upon us...*

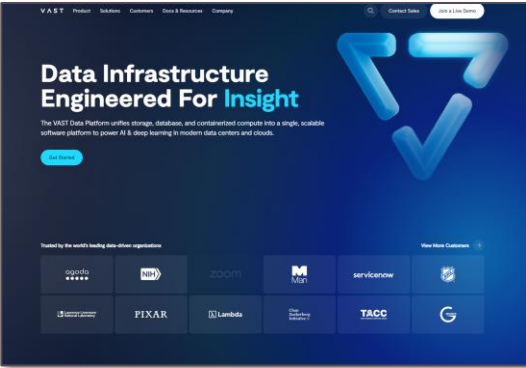
*...We at VAST believe that the path to the greatest potential gain is to simplify and reduce the fundamental challenges that need to be resolved. If we can build a simple approach to encompass nearly all of the infrastructure layers needed for AI, without compromise... customers supremely benefit.*

- VAST Data CEO Renen Hallak, 5/25

Source: VAST Data, Silicon Angle

# AI Monetization...Data Storage / Management / Processing = VAST Data Lifetime Sales From 0 to \$2B in Just Over Six Years

VAST Data – 1/19-5/25, 每 VAST Data



一切都在加速。随着模型构建者在彼此的发现基础上进行构建并不断突破界限，人工智能的进步速度不断提高。虽然自 2022 年初以来我们一直在讨论思维机器，但过去 12 个月中推理模型的出现意味着思维机器的时代实际上已经来临……

… 我们 VAST 认为，获得最大潜在收益的途径是简化并减少需要解决的根本挑战。如果我们能够构建一种简单的方法来涵盖几乎所有人工智能所需的基础设施层，且不妥协……客户将获得极大的好处。

- VAST Data 首席执行官 Renen Hallak, 5/25

Source: VAST Data, Silicon Angle



*AI-Related Cost Ramps Relative to Revenue =*

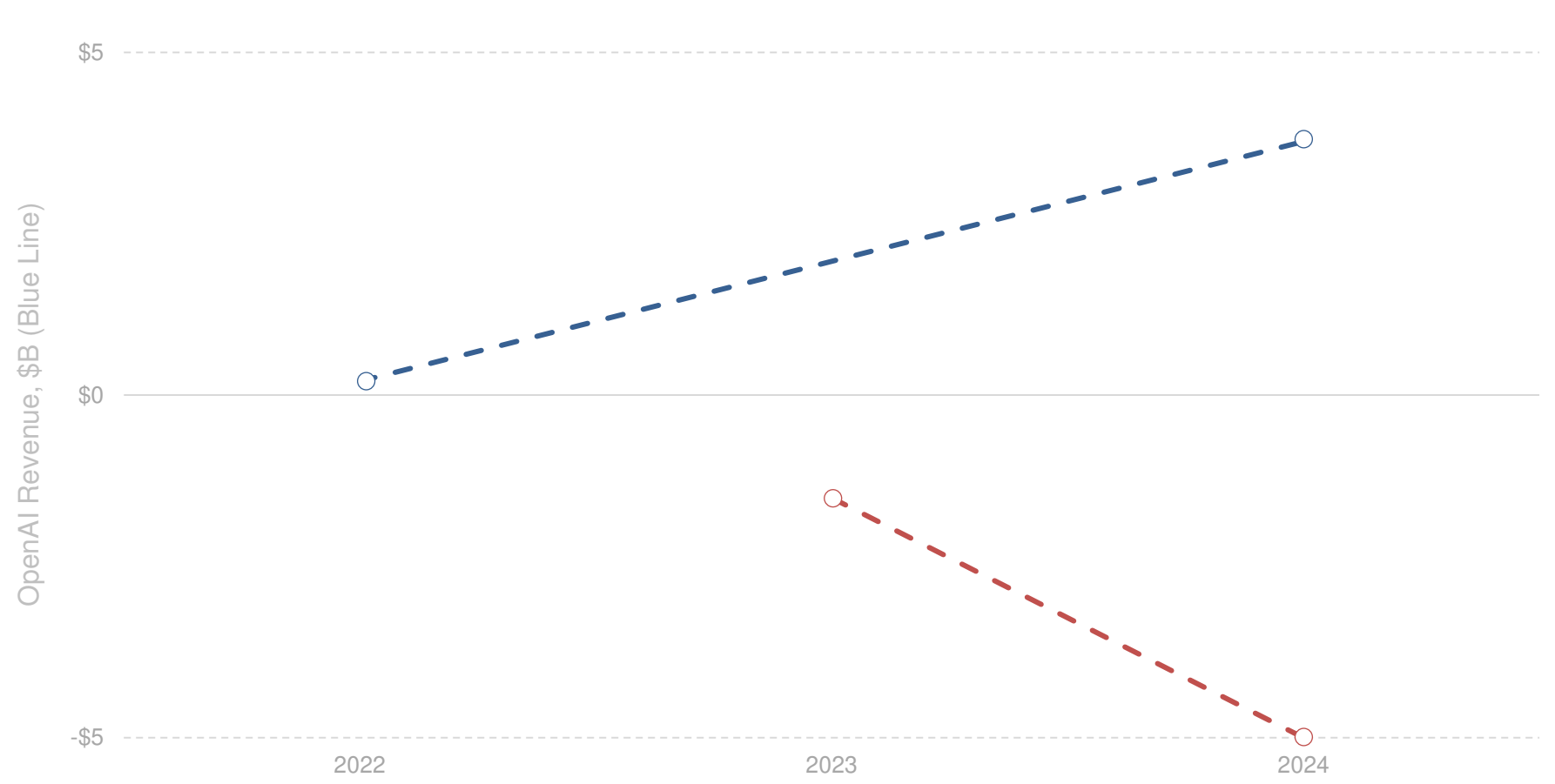
*Can Be Head-Turning*

相对于收入 = 的 AI 相关成本增长

*Can Be Head-Turning*

AI Monetization – OpenAI = Revenue vs. Compute Expense, per *The Information*

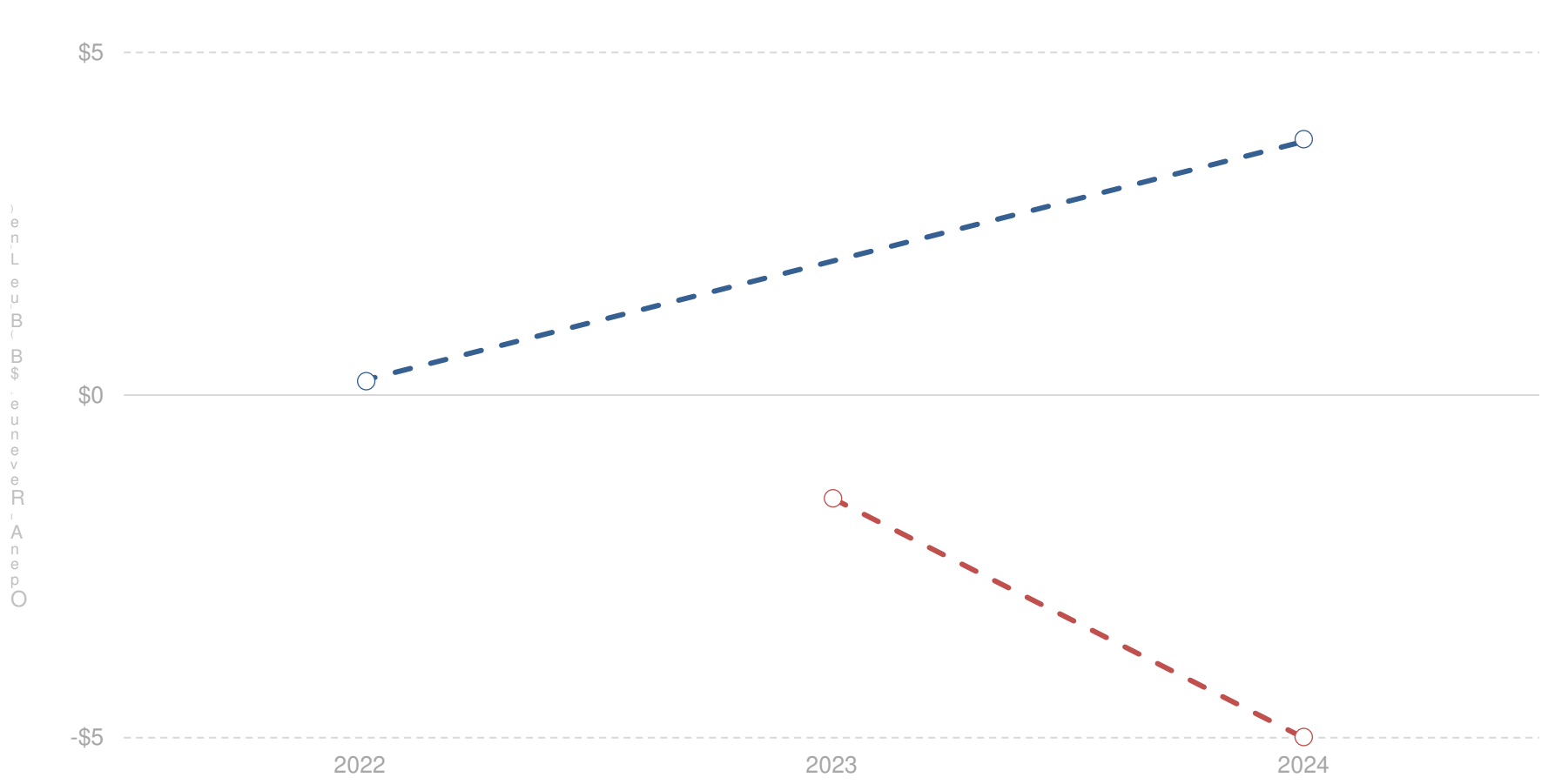
OpenAI Revenue & Compute Expense (\$B) by Year – 2022-2024, per *The Information*



Note: No compute expense data available in 2022. Figures are estimates based off public reports & The Information reporting.  
Source: The Information (4/25 and prior) ([link](#), [link](#), [link](#), [link](#), [link](#) & [link](#))

AI 货币化 – OpenAI = 收入与计算费用, 根据 The Information 报道

OpenAI Revenue & Compute Expense (\$B) by Year – 2022-2024, per *The Information*



注意：2022 年没有可用的计算费用数据。数字是根据公开报告和 The Information 的报告估算的。来源：The Information (4/25 及之前) ([link](#), [link](#), [link](#), [link](#) & [link](#))

AI Monetization – Microsoft / Amazon / Alphabet / Meta = CapEx Up...Free Cash Flow Margins Down

AI 货币化 – Microsoft / Amazon / Alphabet / Meta = 资本支出上升 … 自由现金流利润率下降

Capital Expenditure, Free Cash Flow Margin, Revenue Growth – C2023-C2024, per Capital IQ				
		CapEx	→ Free Cash Flow Margin	vs. Revenue
Microsoft	C2023	\$35B	30%	\$228B
	C2024	\$56B	27%	\$262B
	Y/Y Change	+58%	-10%	+15%
Amazon	C2023	\$53B	6%	\$575B
	C2024	\$83B	5%	\$638B
	Y/Y Change	+57%	-8%	+11%
Alphabet (Google)	C2023	\$32B	23%	\$307B
	C2024	\$52B	21%	\$350B
	Y/Y Change	+63%	-8%	+14%
Meta Platforms (Facebook)	C2023	\$27B	33%	\$135B
	C2024	\$37B	33%	\$165B
	Y/Y Change	+38%	<1%	+22%

资本支出、自由现金流利润率、收入增长 – C2023-C2024, 根据 Capital IQ				
		资本支出	→ 自由现金流利润率	vs. Revenue
Microsoft	C2023	\$35B	30%	\$228B
	C2024	\$56B	27%	\$262B
	同比变化	+58%	-10%	+15%
Amazon	C2023	\$53B	6%	\$575B
	C2024	\$83B	5%	\$638B
	同比变化	+57%	-8%	+11%
Alphabet (Google)	C2023	\$32B	23%	\$307B
	C2024	\$52B	21%	\$350B
	同比变化	+63%	-8%	+14%
Meta Platforms (Facebook)	C2023	\$27B	33%	\$135B
	C2024	\$37B	33%	\$165B
	同比变化	+38%	<1%	+22%

Note: FCF calculated as cash flow from operations less capex to standardize, as only some companies subtract finance leases and Amazon adjusts FCF for gains on sale of equipment. Amazon statistics shown for both AWS & Retail; FCF not broken out across subsidiaries. Source: Capital IQ (5/25)

注意：FCF 计算为经营活动产生的现金流量减去资本支出以进行标准化，因为只有部分公司扣除了融资租赁，而 Amazon 调整了出售设备的收益的 FCF。显示的 Amazon 统计数据包括 AWS 和零售；FCF 未在子公司之间划分。来源：Capital IQ (5/25)

*So...We Have...*

*High Revenue Growth +  
High Cash Burn +  
High Valuations +  
High Investment Levels =*

*Good News for Consumers...  
Others TBD...*

*So...We Have...*

高营收增长 + 高现金  
消耗 + 高估值 + 高投资水  
平 =

对消费者来说是好消息 ..... 其他  
待定 .....

Select Private AI Model Companies – 5/13/25 = ~\$11B+ Annualized Revenue vs. ~\$95B Raised...

Foundation Model Estimated Revenue & Capital Raised – 5/13/25, per Pitchbook, *The Information*, *Bloomberg*, *The Wall Street Journal* & *CNBC*

Company	Annualized Revenue (\$MM)	Total Raised To-Date (\$MM)
OpenAI	9,200 <small>(4/25 estimated)</small>	63,920 <small>(Last Raise: 3/25)</small>
Anthropic	2,000 <small>(3/25)</small>	18,000 <small>(Last Raise: 3/25)</small>
xAI	Materially North of 100* <small>(4/25)</small>	12,130 <small>(Last Raise: 11/24)</small>
Perplexity	120 <small>(5/25)</small>	1,410 <small>(Last Raise: 5/25)</small>

\*Select media reports have xAI revenue being as high as \$1B as of 4/25. Note: OpenAI annualized revenue estimated based upon full-year 2024 & 2025 revenue estimates as published by The Information & Bloomberg, assuming linear revenue growth. Figures are rounded. Source: Source: Pitchbook (5/25), The Information ([link](#)), Bloomberg ([link](#) & [link](#)) & CNBC ([link](#) & [link](#))

Select Private AI Model Companies – 5/13/25 =~\$11B + 年度化收入 vs. ~\$95B Raised...

基础模型预估收入和融资额 – 5/13/25, 数据来源: Pitchbook、 The Information、 Bloomberg、 The Wall Street Journal & CNBC

公司	年化收入（百万美元）	总融资额 aised 至今 (\$MM)
OpenAI	9,200 <small>（ 4/25 估计）</small>	63,920 <small>（上次 R aise: 3/25)</small>
Anthropic	2,000 <small>(3/25)</small>	18,000 <small>（最新一 轮融资 aise: 3/25)</small>
xAI	实质性地 超过 100* <small>(4/25)</small>	12,130 <small>（ 上次融 aise: 11/24)</small>
Perplexity	120 <small>(5/25)</small>	1,410 <small>（最新一 轮融资 aise: 5/25)</small>

\* 选择媒体报道称，截至 4 月 25 日，xAI 的收入高达 10 亿美元。注意：OpenAI 的年度化收入是根据 The Information 和 Bloomberg 发布的 2024 年和 2025 年全年收入估算得出的，假设收入呈线性增长。数字已四舍五入。来源：来源：Pitchbook (5/25)、 The Information ( [链接](#) )、 Bloomberg ( [链接](#) & [链接](#) ) & CNBC ( [链接](#) & [链接](#) )

...Select Private AI Model Companies – 5/13/25 = High Valuation-to-Revenue Multiples

… 选择私有 AI 模型公司 – 5/13/25 = 高估值与收入倍数

Foundation Model Estimated Revenue Multiple – 5/13/25,  
per Pitchbook, *The Information*, *Bloomberg*, *The Wall Street Journal* & *CNBC*

Company	Annualized Revenue (\$MM)	Latest Valuation (\$MM)	Revenue Multiple
OpenAI	9,200 <small>(4/25 estimated)</small>	300,000 <small>(3/25)</small>	33x
Anthropic	2,000 <small>(3/25)</small>	61,500 <small>(3/25)</small>	31x
xAI	Materially North of 100* <small>(4/25)</small>	80,000 <small>(3/25)</small>	N/A
Perplexity	120 <small>(5/25)</small>	9,000 <small>(12/24)</small>	75x

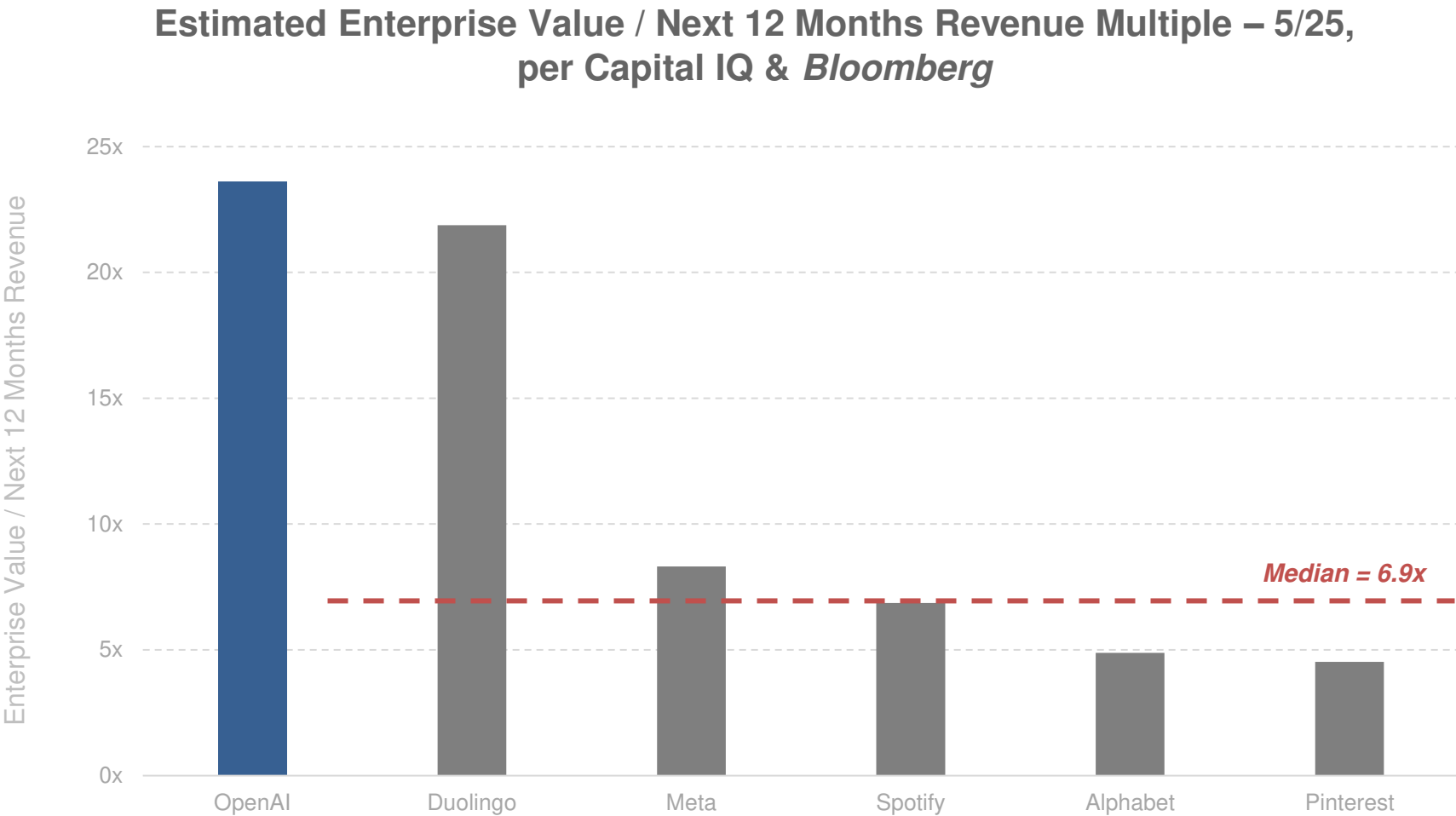
Foundation Model Estimated Revenue Multiple – 5/13/25,  
per Pitchbook, *The Information*, *Bloomberg*, *The Wall Street Journal* & *CNBC*

公司	年度化收入（百万美元）	最新估值（百万美元）	Revenue Multiple
OpenAI	9,200 <small>（ 4/25 估计）</small>	300,000 <small>（ 3/25 ）</small>	33x
Anthropic	2,000 <small>（ 3/25 ）</small>	61,500 <small>（ 3/25 ）</small>	31x
xAI	Materially 高于 100* <small>（ 4/25 ）</small>	80,000 <small>(3/25)</small>	N/A
Perplexity	120 <small>（ 5/25 ）</small>	9,000 <small>(12/24)</small>	75x

*\*Select media reports have xAI revenue being as high as \$1B as of 4/25. Note: OpenAI annualized revenue estimated based upon full-year 2024 & 2025 revenue estimates as published by The Information & Bloomberg, assuming linear revenue growth. xAI valuation per Elon Musk. Figures are rounded. Perplexity was reported to be in advanced talks to raise capital at a \$14B post-money valuation as of 5/14/25; however, as this is not finalized at time of publication, we quote their last finalized funding round here. Source: Pitchbook (5/25), The Information ([link](#)), Bloomberg ([link](#) & [link](#)) & CNBC ([link](#) & [link](#))*

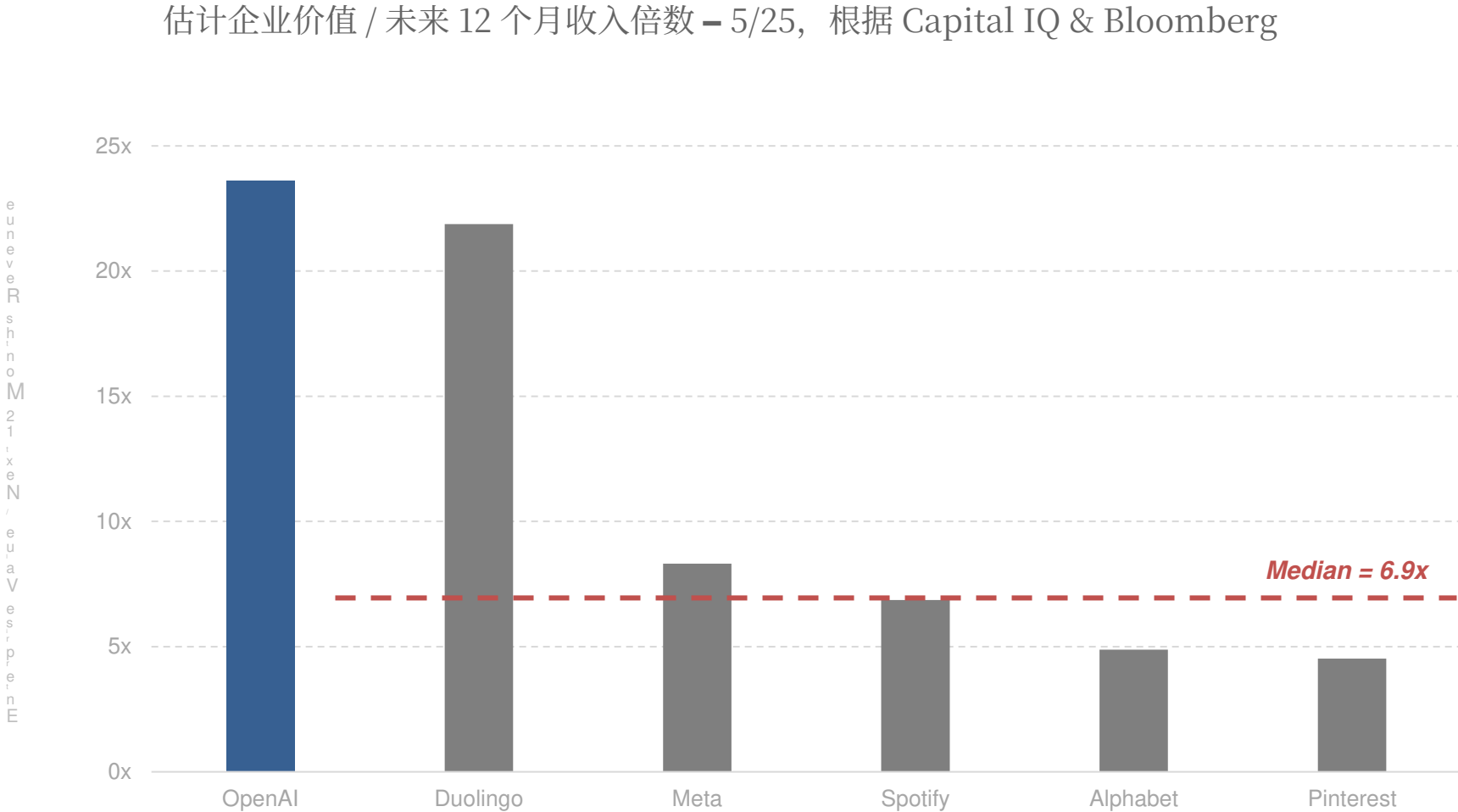
*\* 据一些媒体报道，截至 4 月 25 日，xAI 的收入高达 10 亿美元。注：OpenAI 的年度收入是根据 The Information 和彭博社发布的 2024 和 2025 全年收入估算值估算的，假设收入呈线性增长。xAI 的估值来自埃隆·马斯克。数字已四舍五入。据报道，Perplexity 正在就以 140 亿美元的投后估值筹集资金进行深入谈判，截至 2025 年 5 月 14 日；但是，由于在发布时尚未最终确定，因此我们在此引用他们上次最终确定的融资轮次。来源：Pitchbook (5/25)、The Information ( [链接](#) )、彭博社 ( [链接](#) & [链接](#) ) 和 CNBC ( [链接](#) & [链接](#) )*

Valuation-to-Revenue Multiple – OpenAI = Looks Expensive...



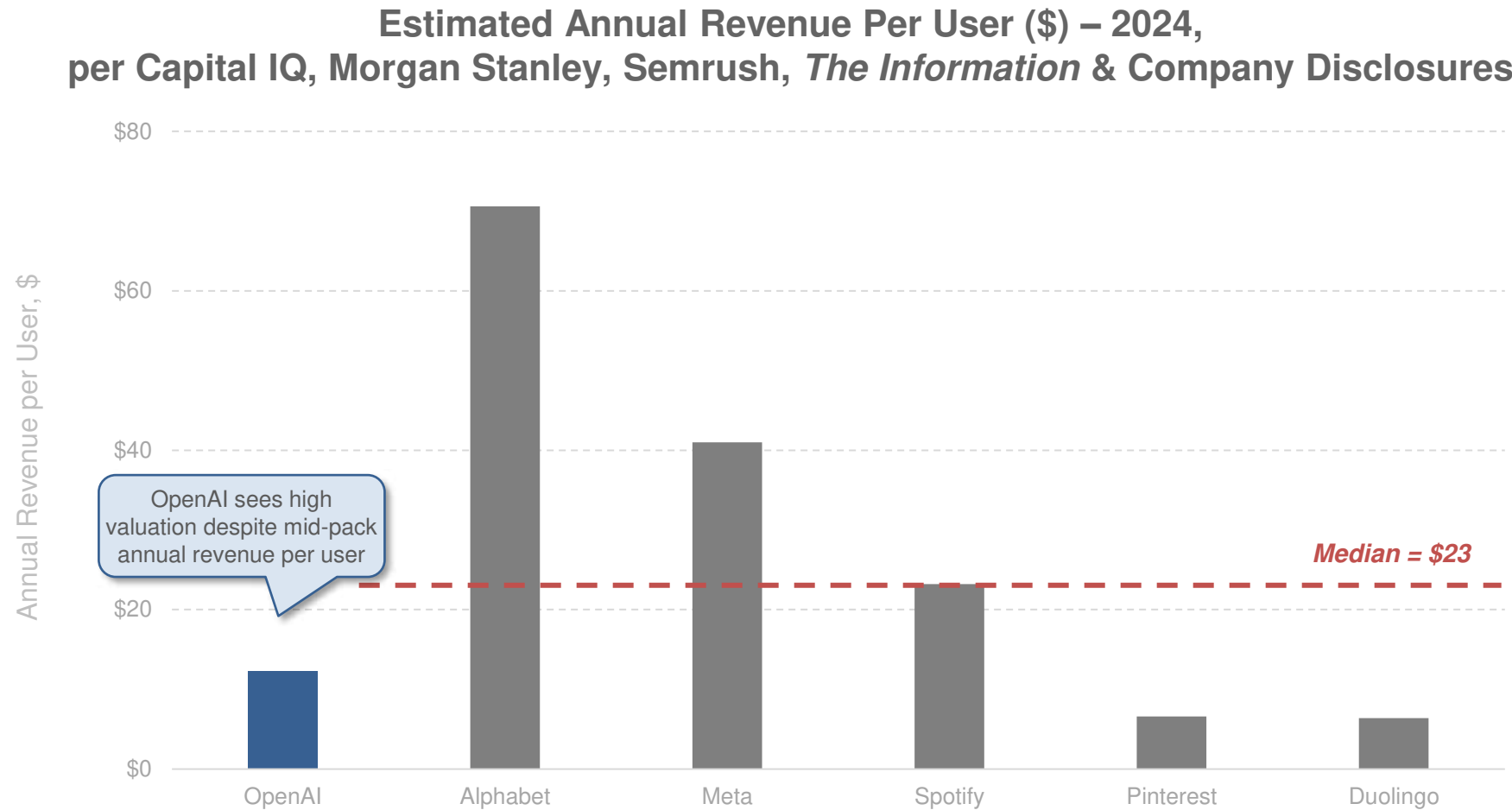
Note: OpenAI figures are estimates. Next 12 months revenue multiples for companies other than OpenAI are consensus estimates per Capital IQ. OpenAI NTM revenue estimates are as of 12/24 due to data availability. Source: Capital IQ (5/15/25), Bloomberg (link)

Valuation-to-Revenue Multiple – OpenAI = Looks Expensive...



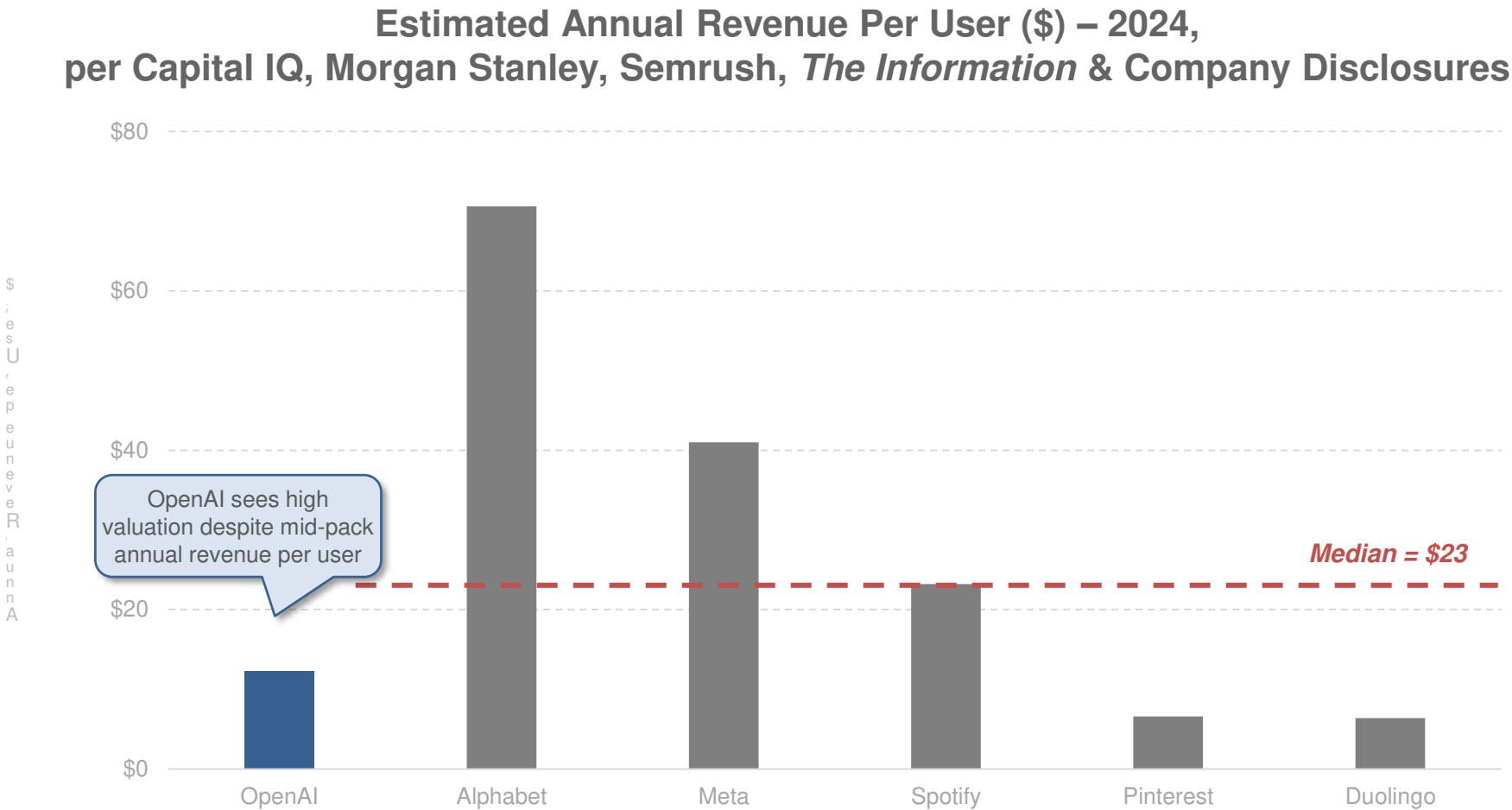
注意：OpenAI 的数据为估计值。除 OpenAI 之外的公司的未来 12 个月收入倍数是根据 Capital IQ 得出的共识估计值。由于数据可用性，OpenAI NTM 收入估计截至 12/24。来源：Capital IQ (5/15/25)，Bloomberg (link)

...Revenue-per-User Multiple – OpenAI = In-the-Range



Note: OpenAI figures are estimates as of 4/25. All other public-company figures are as of 12/31/24, using CY2024 data. OpenAI data uses WAUs due to data availability (conservatively assumed as MAUs); other figures use MAUs. Here we assume average weekly active ChatGPT users of 300MM based off OpenAI's 12/24 disclosure. We estimate 2024 ChatGPT revenue of \$3.7B, per company estimates. Monthly active user figures are estimates for Alphabet based off website traffic measurements & global internet user data. Meta last reported MAPs for app family in Q4:23, we conservatively assume no growth since. Source: Capital IQ (12/24), The Information (4/25 and prior) ([link](#), [link](#), [link](#), [link](#) & [link](#)), Semrush (11/24), Morgan Stanley, ITU, company disclosures, BOND estimates

...Revenue-per-User Multiple – OpenAI = In-the-Range



注意：OpenAI 的数据为 4 月 25 日的估计值。所有其他上市公司的数字均为 2024 年 12 月 31 日的数字，使用 CY2024 数据。由于数据可用性，OpenAI 数据使用 WAUs（保守地假定为 MAUs）；其他数字使用 MAUs。在此，我们根据 OpenAI 的 12/24 披露，假设 ChatGPT 的平均每周活跃用户为 3 亿。根据公司估计，我们估计 2024 年 ChatGPT 收入为 37 亿美元。Alphabet 的月活跃用户数据是根据网站流量测量和全球互联网用户数据估算得出的。Meta 上次报告应用系列 MAPs 的时间是 23 年第 4 季度，我们保守地假设此后没有增长。来源：Capital IQ (12/24)、The Information（4/25 及之前）（[链接](#)、[链接](#)、[链接](#)、[链接](#)和[链接](#)）、Semrush (11/24)、摩根士丹利、ITU、公司披露、BOND 估计



As global digital user bases have grown and potential rapidity of usage traction has risen in tandem, areas of corporate investment (for companies new and old) have become increasingly competitive and capital-intensive. The AI tech cycle of creative disruption has historical analogs.

Head turners of the semi-recent past include Apple’s near bankruptcy in 1997 when its market capitalization was \$1.7B\*, now \$3.2T.

Amazon.com’s near death moment happened in Q4:00 when it reported a net loss of -\$545MM on revenue of \$972MM.

Founder and then-CEO Jeff Bezos noted in the 2000 Shareholder Report that *It’s been a brutal year for many in the capital markets and certainly for Amazon.com shareholders. As of this writing, our shares are down more than 80% from when I wrote you last year.* At post-loss trough in Q3:01 its market cap was \$2.2B while it supported 23MM active customer accounts. The market cap is now \$2.2T.

All in, Amazon lost -\$3B in the twenty-seven quarters between its launch in Q2:97 and the end of its first net income-positive year (2003). For its most recent twenty-seven most recent quarters (Q3:18-Q1:25), Amazon’s cumulative net income was \$176B.

Google’s IPO filing (April 2004) noted that in Q1:04, after having only raised a Series A funding round, it spent 22% of revenue (\$86MM of \$390MM) on capital expenditures – at the time it was an incomprehensibly high number. It went public at a \$23B market cap, now \$2.0T...

*\*Market capitalization taken as of 7/1/97. Microsoft finalized its investment in Apple just over one month later, on 8/6/97. Note: Present market capitalization figures are shown as of 5/14/25.*

As global digital user bases have grown and potential rapidity of usage traction has risen in tandem, areas of corporate investment (for companies new and old) have become increasingly competitive and capital-intensive. The AI tech cycle of creative disruption has historical analogs.

Head turners of the semi-recent past include Apple’s near bankruptcy in 1997 when its market capitalization was \$1.7B\*, now \$3.2T.

Amazon.com’s near death moment happened in Q4:00 when it reported a net loss of -\$545MM on revenue of \$972MM.

Founder and then-CEO Jeff Bezos noted in the 2000 Shareholder Report that *It’s been a brutal year for many in the capital markets and certainly for Amazon.com shareholders. As of this writing, our shares are down more than 80% from when I wrote you last year.* At post-loss trough in Q3:01 its market cap was \$2.2B while it supported 23MM active customer accounts. The market cap is now \$2.2T.

All in, Amazon lost -\$3B in the twenty-seven quarters between its launch in Q2:97 and the end of its first net income-positive year (2003). For its most recent twenty-seven most recent quarters (Q3:18-Q1:25), Amazon’s cumulative net income was \$176B.

Google’s IPO filing (April 2004) noted that in Q1:04, after having only raised a Series A funding round, it spent 22% of revenue (\$86MM of \$390MM) on capital expenditures – at the time it was an incomprehensibly high number. It went public at a \$23B market cap, now \$2.0T...

*\* 市值截至 1997 年 7 月 1 日。微软在一个多月后的 1997 年 8 月 6 日完成了对苹果的投资。注意：目前的市值数据截至 25 年 5 月 14 日。*

...Uber burned -\$17B\* between 2016 and 2022 (and materially more before that) before its first free cash flow-positive year in 2023.  
In 2022, it had 131MM monthly active platform consumers.  
Uber’s last equity financing was a Series G.  
Its fully-diluted IPO market cap was \$82B, now \$189B.

Tesla burned -\$9.2B between 2009 and 2018 before becoming free cash flow positive in 2019.  
In the ten years between 2009 and 2018, it lost a cumulative -\$5.6B delivering ~540K vehicles.  
It went public in 2010 at a market cap of \$1.6B.  
From 2019-2024, it then earned \$40B delivering 6.7MM vehicles.  
Its market cap is now \$1.1T.

It is important to remember – most of the time, when all is said and done – a business’s valuation should represent the present value of its future free cash flows.  
The aforementioned companies – with aggressive cash burn – tested this premise hard, built large-scale data-driven network effects based on product excellence / constant improvement, developed technology-driven competitive advantage and ultimately proved the naysayers wrong.

Only time will tell which side of the money-making equation the current AI aspirants will land.

...Uber 烧掉了 -\$17B\* 在 2016 年至 2022 年之间（并且在此之前更多），然后在 2023 年实现了第一个自由现金流为正的年份。2022 年，它拥有 1.31 亿月活跃平台消费者。Uber 的最后一轮股权融资是 G 轮。其完全稀释后的首次公开募股（IPO）市值为 820 亿美元，现在为 1890 亿美元。

特斯拉在 2009 年至 2018 年间烧掉了 -92 亿美元，然后在 2019 年实现了自由现金流为正。在 2009 年至 2018 年之间的十年中，它累计亏损了 -56 亿美元，交付了 ~540K 辆汽车。它于 2010 年上市，市值为 16 亿美元。从 2019-2024 年，它通过交付 670 万辆汽车赚取了 400 亿美元。其市值现在为 1.1 万亿美元。

重要的是要记住 – 大多数时候，当一切尘埃落定 – 一家企业的估值应该代表其未来自由现金流的现值。上述公司 – 凭借激进的现金消耗 –，对这一前提进行了严峻的考验，基于卓越的产品 / 持续改进，建立了大规模的、数据驱动的网络效应，开发了技术驱动的竞争优势，并最终证明了那些唱反调的人是错误的。

只有时间才能证明当前的人工智能有志者最终会落在赚钱等式的哪一边。

*\*Measured as unlevered free cash flow.  
Note: Present market capitalization figures are shown as of 5/14/25.*

*\* 以非杠杆自由现金流衡量。注意：目前的市值数据截至 5/14/25。*

*Usage + Cost + Loss Growth =*

*Unprecedented...*

*What About **Future** Monetization + Profits?*

使用量 + 成本 + 损失增长 =

前所未有 ……

未来的货币化 + 利润如何?

AI Monetization Possibilities =  
New Entrants & / Or Tech Incumbents?

AI 货币化可能性 =  
新的进入者和 / 或科技巨头?

To understand where AI model economics may be heading, one can look at the mounting tension between capabilities and costs.

Training the most powerful large language models (LLMs) has become one of the most expensive / capital-intensive efforts in human history. As the frontier of performance pushes toward ever-larger parameter counts and more complex architectures, model training costs are rising into the billions of dollars.

Ironically, this race to build the most capable general-purpose models may be accelerating commoditization and driving diminishing returns, as output quality converges across players and differentiation becomes harder to sustain. At the same time, the cost of applying/using these models – known as inference – is falling quickly. Hardware is improving – for example, NVIDIA’s 2024 Blackwell GPU consumes 105,000x less energy per token than its 2014 Kepler GPU predecessor. Couple that with breakthroughs in models’ algorithmic efficiency, and the cost of inference is plummeting.

Inference represents a new cost curve, and – unlike training costs – it’s arcing down, not up. As inference becomes cheaper and more efficient, the competitive pressure amongst LLM providers increases – not on accuracy alone, but also on latency, uptime, and cost-per-token\*. What used to cost dollars can now cost pennies. And what cost pennies may soon cost fractions of a cent...

*\*Cost-per-token = The expense incurred for processing or generating a single token (a word, sub-word, or character) during the operation of a language model. It is a key metric used to evaluate the computational efficiency and cost-effectiveness of deploying AI models, particularly in applications like natural language processing.*

要了解 AI 模型经济的未来走向，可以关注能力与成本之间日益加剧的紧张关系。

训练最强大的大型语言模型 (LLM) 已成为人类历史上最昂贵 / 资本密集的努力之一。随着性能前沿向着越来越大的参数数量和更复杂的架构推进，模型训练成本正在上升到数十亿美元。

具有讽刺意味的是，这场构建最强大的通用模型的竞赛可能会加速商品化并导致收益递减，因为各个参与者的输出质量趋于一致，并且差异化变得难以维持。与此同时，应用 / 使用这些模型的成本 —— 被称为推理 —— 正在迅速下降。硬件正在改进 —— 例如，NVIDIA 的 2024 Blackwell GPU 的每 token 功耗比 2014 Kepler GPU 前代产品低 105,000 倍。再加上模型算法效率的突破，推理成本正在暴跌。

推理代表了一条新的成本曲线，并且 – 与训练成本不同 – 它的弧线向下，而不是向上。随着推理变得更便宜和更有效，LLM 提供商之间的竞争压力增加 – 不仅在准确性方面，还在延迟、正常运行时间和每个 token 的成本方面 \*。过去需要花费几美元的东西现在可能只需要几美分。而过去需要几美分的东西可能很快只需要几分之一美分 …

*\*Cost-per-token = The expense incurred for processing or generating a single token (a word, sub-word, or character) during the operation of a language model. It is a key metric used to evaluate the computational efficiency and cost-effectiveness of deploying AI models, particularly in applications like natural language processing.*

...The implications are still unfolding. For users (and developers), this shift is a gift: dramatically lower unit costs to access powerful AI. And as end-user costs decline, creation of new products and services is flourishing, and user and usage adoption is rising.

For model providers, however, this raises real questions about monetization and profits. Training is expensive, serving is getting cheap, and pricing power is slipping. The business model is in flux. And there are new questions about the one-size-fits-all LLM approach, with smaller, cheaper models trained for custom use cases\* now emerging.

Additionally, traditional business moats are being disrupted. Look no further than Google. The company launched AI Overviews in May of last year – they sit above many Google search results. The company highlighted it had 1.5B AI Overviews MAUs as of 4/25...it’s notable that in the last few weeks, Google began adding advertisements to select AI Overviews.

Will providers try to build horizontal platforms? Will they dive into specialized applications? Will one or two leaders drive dominant user and usage share and related monetization, be it subscriptions (easily enabled by digital payment providers), digital services, ads, etc.? Only time will tell. In the short term, it’s hard to ignore that the economics of general-purpose LLMs look like commodity businesses with venture-scale burn.

*\*E.g., OpenEvidence*

...The implications are still unfolding. For users (and developers), this shift is a gift: 大幅降低了访问强大 AI 的单位成本。随着最终用户成本的下降，新产品和服务的创建蓬勃发展，用户和使用采用率也在上升。

然而，对于模型提供商来说，这引发了关于货币化和利润的实际问题。训练成本很高，服务成本越来越低，定价权正在下降。商业模式正在变化。关于一刀切的 LLM 方法也出现了新的问题，针对自定义用例 \* 训练的更小、更便宜的模型现在正在涌现。

此外，传统的商业护城河正在被打破。看看谷歌就知道了。该公司去年 5 月推出了 AI Overviews – 它们位于许多 Google 搜索结果之上。该公司强调，截至 4 月 25 日，其 AI Overviews 的 MAU 为 15 亿 …… 值得注意的是，在过去几周，谷歌开始向精选的 AI Overviews 添加广告。

供应商会尝试构建水平平台吗？他们会投入到专门的应用程序中吗？是否会有一两个领导者推动主要的用户和使用份额以及相关的货币化，无论是订阅（数字支付提供商可以轻松实现）、数字服务、广告等？时间会证明一切。短期内，很难忽视通用 LLM 的经济效益看起来像是具有风险规模消耗的商品业务。

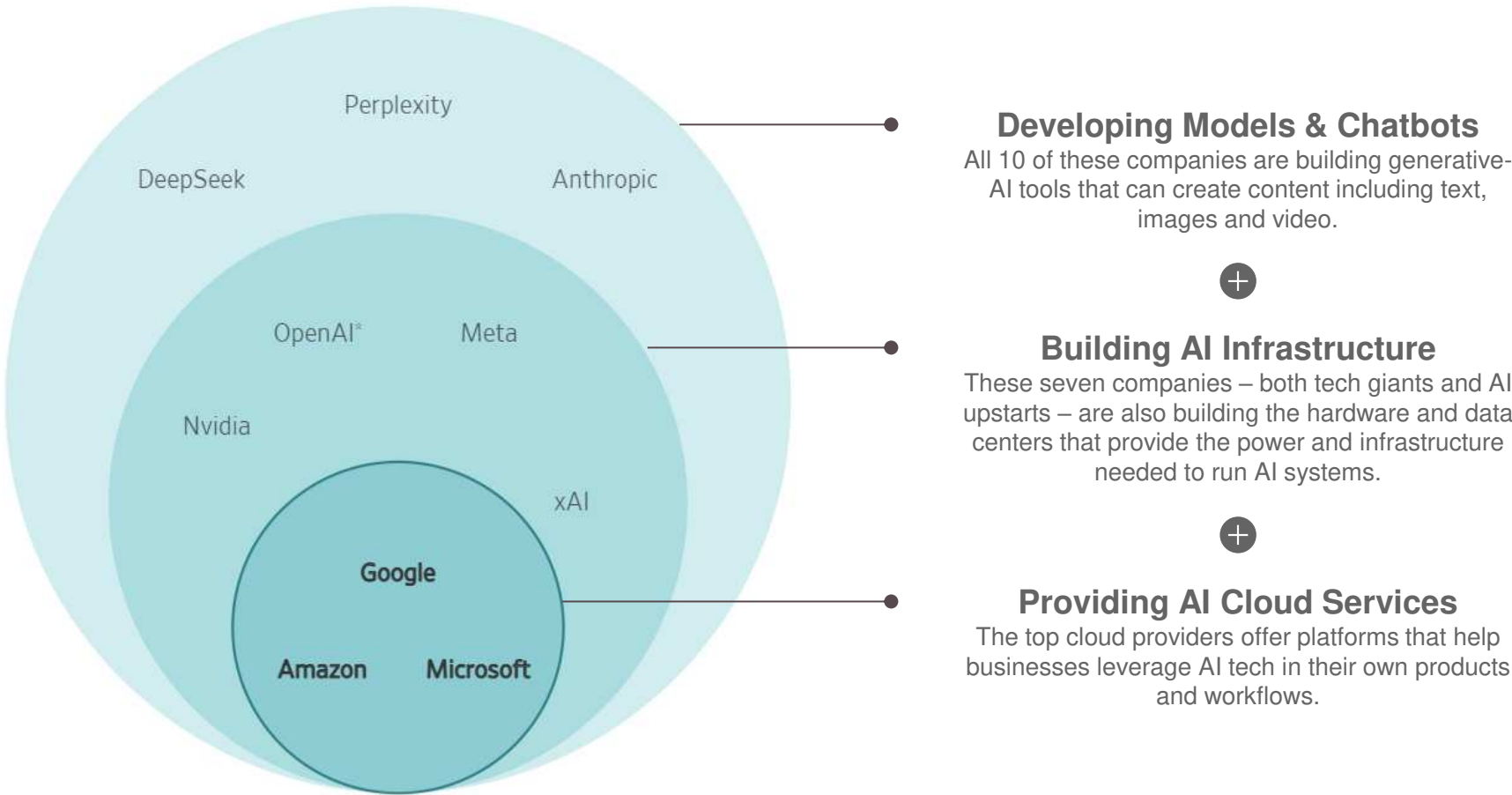
*\* 例如，OpenEvidence*



# AI Company Landscape = Varying Degrees of Vertical Integration

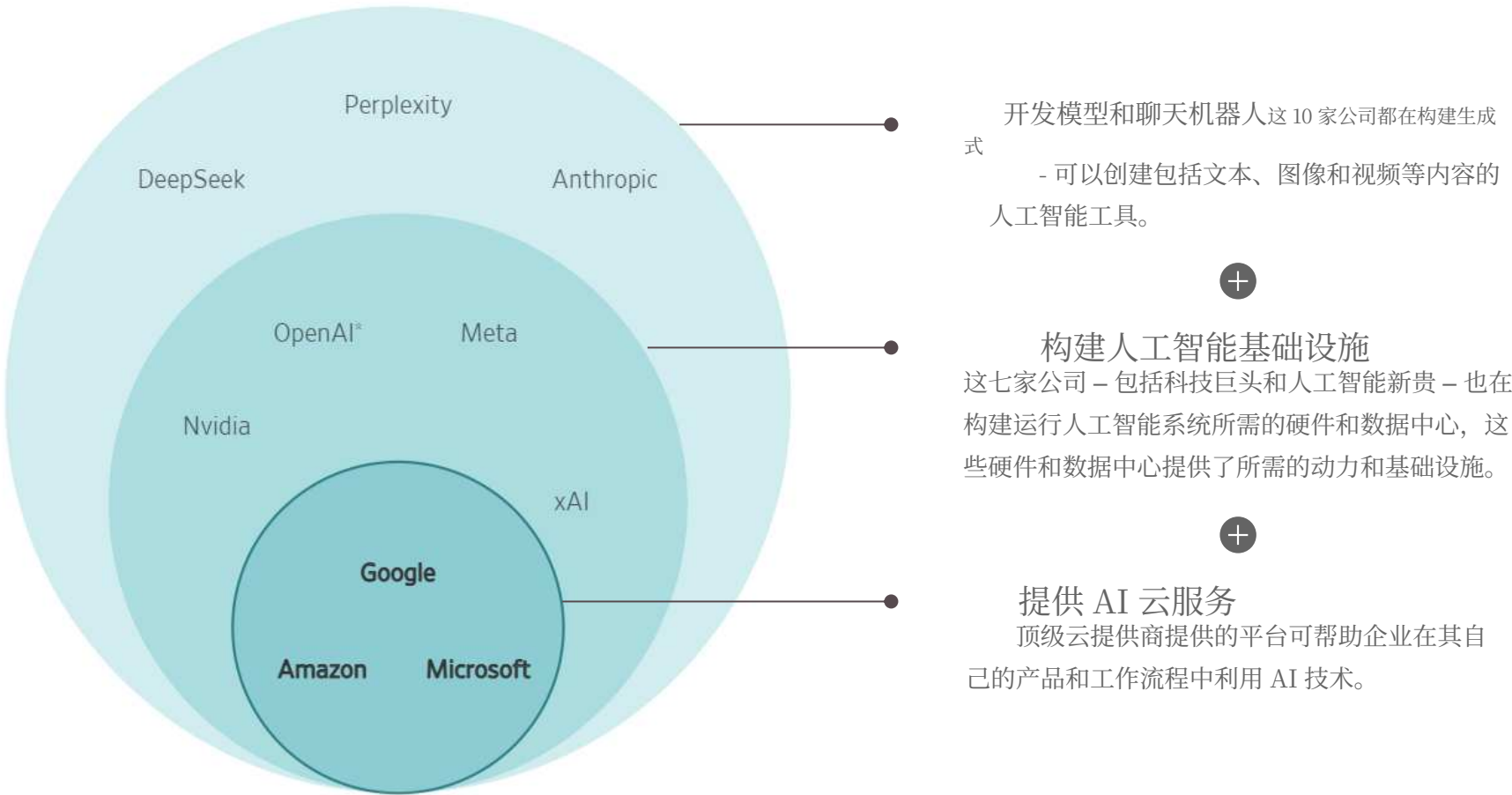
# AI 公司格局 = 不同程度的垂直整合

Specializations of Ten Leading AI Companies – 4/25, per *The Wall Street Journal*



\*Has a partnership with Oracle, SoftBank and MGX to build out the proposed Stargate data-center network.  
Source: Wall Street Journal, 'Here's How Big the AI Revolution Really Is, in Four Charts' (4/25)

Specializations of Ten Leading AI Companies – 4/25, per *The Wall Street Journal*



\* 与 Oracle、软银和 MGX 合作，构建拟议的 Stargate 数据中心网络。来源：华尔街日报，“人工智能革命的规模究竟有多大，见四张图表”（4/25）

AI Monetization Possibilities =

**New Entrants** & / Or Tech Incumbents?

AI 货币化可能性 =

新的进入者和 / 或科技巨头?



*AI – New Entrants =*

*Rapidly Laying Groundwork*

人工智能 – 新进入者 =

迅速奠定基础

AI Monetization...Foundation Models = Consumer Subscription Models Driving Monetization...

AI 货币化 ... 基础模型 = 消费者订阅模型驱动货币化 ...

OpenAI ChatGPT, xAI Grok, Google Gemini, Anthropic Claude & Perplexity Consumer Pricing – 5/25, per Companies

OpenAI ChatGPT, xAI Grok, Google Gemini, Anthropic Claude & Perplexity 消费者定价 – 5/25, 每家公司

Free

Explore how AI can help with everyday tasks

- ✓ Access to GPT-4o mini
- ✓ Real-time data from the web with search
- ✓ Limited access to GPT-4o, OpenAI o1-mini, and deep research
- ✓ Limited access to file uploads, data analysis, image generation, and voice mode
- ✓ Code edits with the ChatGPT developer app for macOS
- ✓ Use custom GPTs

Have an existing plan? See [pricing info](#)

\$0 / month

Get Free

Plus

Level up productivity and creativity with expanded access

- ✓ Everything in Free
- ✓ Extended limits on messaging, file uploads, data analysis, and image generation
- ✓ Standard and advanced voice mode with video and screen sharing
- ✓ Access to deep research and multiple reasoning models (OpenAI o1, OpenAI o1-mini, and OpenAI o1-mini-high)
- ✓ Access to a research preview of GPT-4o, our largest model yet
- ✓ Create and use projects, teams, and custom GPTs
- ✓ Opportunities to test new features

\$20 / month

Get Plus

Limits apply >

Pro

Get the best of OpenAI with the highest level of access

- ✓ Everything in Plus
- ✓ Unlimited access to all reasoning models and GPT-4o
- ✓ Unlimited access to advanced voice with higher limits for video and screen sharing
- ✓ Access to OpenAI o1-pro models, which saves more compute for the best answers to the hardest questions
- ✓ Extended access to deep research
- ✓ Extended access to file editor generation
- ✓ Access to a research preview of OpenAI o1, subject to change without notice

\$200 / month

Get Pro

OpenAI ChatGPT  
\$0 (Free) / \$20 (Plus) / \$200 (Pro)  
per Month

About X Premium

Subscribe today

3 Essential price updates

What is X Premium? X Premium is an opt-in, paid subscription service that elevates quality conversations on the platform. X Premium has three tiers: Basic, Premium, and Premium+, with more features available in each higher tier.

Basic: Includes essential Premium features like editing posts, longer posts and longer video uploads, reply prioritization, best formatting, business features, custom app icons, and more.

Premium: Includes all Basic features plus a checklist, reduced ads, access to apply for creator revenue sharing and creator subscriptions, larger reply prioritization, ID verification, Media Studio, and increased usage with on [Grok](#).

Premium+: Includes all Premium features with additional benefits like higher limits on Grok, no ads anywhere on X, largest app prioritization, faster Search, and Articles. Occasional promoted content may appear.

The complete list of the features is [here](#).

Subscribe now with localized pricing starting at \$3/month or \$32/year plus any tax, e.g. VAT, and your payment method must be on Web in available countries. Click [here](#) for pricing information.

We've also launched a new service called Verified Organizations, which is a service for businesses, governments and experts in X that comes with a gold or platinum badge, VIP support, and more features for organizations. Learn more about Verified Organizations [here](#).

WITH A BUSINESS ACCOUNT

\$0/month

✓ 10 TB of storage

✓ Help with writing, planning, learning and more with Grok

✓ Reasoning by your prompt (not a general model) and writing assistant

WITH A PREMIUM PLAN

\$9.99/month \$0 for the first month

✓ 1 TB of storage

✓ Help with writing, planning, learning and more with Grok

✓ Grok Advanced with access to our most capable models

✓ Choose AI integration across Google apps with Google Assistant, Docs, and more

✓ Priority access to future innovations

✓ Increased ad spend with higher usage limits and premium features

✓ Choose high-quality content with best-in-class other generative AI models, in Google Assistant and Grok

✓ Add more premium benefits across Google

xAI Grok  
\$0 (Free) / \$3 (Basic) / \$8 (Premium) / \$40 (Premium+) per Month<sup>1</sup>

Get the best of Google AI, including Gemini Advanced and 2 TB of storage

Try for 1 month or upgrade

Gemini Advanced is free for students through June 2024

With a business account

\$0/month

✓ 10 TB of storage

✓ Help with writing, planning, learning and more with Gemini

✓ Reasoning by your prompt (not a general model) and writing assistant

WITH A PREMIUM PLAN

\$9.99/month \$0 for the first month

✓ 1 TB of storage

✓ Help with writing, planning, learning and more with Gemini

✓ Gemini Advanced with access to our most capable models

✓ Choose AI integration across Google apps with Google Assistant, Docs, and more

✓ Priority access to future innovations

✓ Increased ad spend with higher usage limits and premium features

✓ Choose high-quality content with best-in-class other generative AI models, in Google Assistant and Grok

✓ Add more premium benefits across Google

Google Gemini  
\$0 (Free) / \$19.99 (AI Pro) / \$250 (AI Ultra) per Month

Free

Explore how AI can help with everyday tasks

- ✓ Access to GPT-4o mini
- ✓ Real-time data from the web with search
- ✓ Limited access to GPT-4o, OpenAI o1-mini, and deep research
- ✓ Limited access to file uploads, data analysis, image generation, and voice mode
- ✓ Code edits with the ChatGPT developer app for macOS
- ✓ Use custom GPTs

Have an existing plan? See [pricing info](#)

\$0 / month

Get Free

Plus

Level up productivity and creativity with expanded access

- ✓ Everything in Free
- ✓ Extended limits on messaging, file uploads, data analysis, and image generation
- ✓ Standard and advanced voice mode with video and screen sharing
- ✓ Access to deep research and multiple reasoning models (OpenAI o1, OpenAI o1-mini, and OpenAI o1-mini-high)
- ✓ Access to a research preview of GPT-4o, our largest model yet
- ✓ Create and use projects, teams, and custom GPTs
- ✓ Opportunities to test new features

\$20 / month

Get Plus

Limits apply >

Pro

Get the best of OpenAI with the highest level of access

- ✓ Everything in Plus
- ✓ Unlimited access to all reasoning models and GPT-4o
- ✓ Unlimited access to advanced voice with higher limits for video and screen sharing
- ✓ Access to OpenAI o1-pro models, which saves more compute for the best answers to the hardest questions
- ✓ Extended access to deep research
- ✓ Extended access to file editor generation
- ✓ Access to a research preview of OpenAI o1, subject to change without notice

\$200 / month

Get Pro

OpenAI ChatGPT  
\$0 (Free) / \$20 (Plus) / \$200 (Pro)  
per Month

About X Premium

Subscribe today

3 Essential price updates

What is X Premium? X Premium is an opt-in, paid subscription service that elevates quality conversations on the platform. X Premium has three tiers: Basic, Premium, and Premium+, with more features available in each higher tier.

Basic: Includes essential Premium features like editing posts, longer posts and longer video uploads, reply prioritization, best formatting, business features, custom app icons, and more.

Premium: Includes all Basic features plus a checklist, reduced ads, access to apply for creator revenue sharing and creator subscriptions, larger reply prioritization, ID verification, Media Studio, and increased usage with on [Grok](#).

Premium+: Includes all Premium features with additional benefits like higher limits on Grok, no ads anywhere on X, largest app prioritization, faster Search, and Articles. Occasional promoted content may appear.

The complete list of the features is [here](#).

Subscribe now with localized pricing starting at \$3/month or \$32/year plus any tax, e.g. VAT, and your payment method must be on Web in available countries. Click [here](#) for pricing information.

We've also launched a new service called Verified Organizations, which is a service for businesses, governments and experts in X that comes with a gold or platinum badge, VIP support, and more features for organizations. Learn more about Verified Organizations [here](#).

WITH A BUSINESS ACCOUNT

\$0/month

✓ 10 TB of storage

✓ Help with writing, planning, learning and more with Gemini

✓ Reasoning by your prompt (not a general model) and writing assistant

WITH A PREMIUM PLAN

\$9.99/month \$0 for the first month

✓ 1 TB of storage

✓ Help with writing, planning, learning and more with Gemini

✓ Gemini Advanced with access to our most capable models

✓ Choose AI integration across Google apps with Google Assistant, Docs, and more

✓ Priority access to future innovations

✓ Increased ad spend with higher usage limits and premium features

✓ Choose high-quality content with best-in-class other generative AI models, in Google Assistant and Grok

✓ Add more premium benefits across Google

xAI Grok  
\$0 (免费) / \$3 (基础版) / 每月 \$40 (高级版 + )

Get the best of Google AI, including Gemini Advanced and 2 TB of storage

Try for 1 month or upgrade

Gemini Advanced is free for students through June 2024

With a business account

\$0/month

✓ 10 TB of storage

✓ Help with writing, planning, learning and more with Gemini

✓ Reasoning by your prompt (not a general model) and writing assistant

WITH A PREMIUM PLAN

\$9.99/month \$0 for the first month

✓ 1 TB of storage

✓ Help with writing, planning, learning and more with Gemini

✓ Gemini Advanced with access to our most capable models

✓ Choose AI integration across Google apps with Google Assistant, Docs, and more

✓ Priority access to future innovations

✓ Increased ad spend with higher usage limits and premium features

✓ Choose high-quality content with best-in-class other generative AI models, in Google Assistant and Grok

✓ Add more premium benefits across Google

Google Gemini  
\$0 (免费) / \$19.99 (AI Pro) / \$250 (AI Ultra) per Month

Free

Try Claude

\$0

Free for everyone

Try Claude

Pro

For everyday productivity

\$17

Per month with annual subscription discount; \$200 billed up front; \$20 if billed monthly

Try Claude

Max

Get the most out of Claude

From \$100

Per person billed monthly

Try Claude

Everything in Free, plus:

- More usage
- Access to unlimited Projects to organize chats and documents
- Ability to search the web
- Extended thinking for complex work
- Connect Google Workspace: email, calendar, and docs
- Ability to use more Claude models

Everything in Pro, plus:

- Choose 5x or 20x more usage than Pro
- Higher output limits for all tasks
- Access Claude Code directly in your terminal
- Access to advanced Research
- Connect any content or tool through integrations
- Early access to advanced Claude features
- Priority access at high traffic times

Anthropic Claude  
\$0 (Free) / \$17<sup>2</sup> (Plus) / \$100 (Max)  
per Month

Give your team their time back

Whether you're a small business or a large enterprise, Perplexity Enterprise Pro enables your team to think strategically rather than search endlessly.

Get started

perplexity Free

perplexity Pro \$20/month

perplexity ENTERPRISE Pro \$40/month per seat or \$4000/year (quote 2024)

Get started for free

Get started

Get started

Perplexity  
\$0 (Free) / \$20 (Pro)  
per Month

Free

Try Claude

\$0

Free for everyone

Try Claude

Pro

For everyday productivity

\$17

Per month with annual subscription discount; \$200 billed up front; \$20 if billed monthly

Try Claude

Max

Get the most out of Claude

From \$100

Per person billed monthly

Try Claude

Everything in Free, plus:

- More usage
- Access to unlimited Projects to organize chats and documents
- Ability to search the web
- Extended thinking for complex work
- Connect Google Workspace: email, calendar, and docs
- Ability to use more Claude models

Everything in Pro, plus:

- Choose 5x or 20x more usage than Pro
- Higher output limits for all tasks
- Access Claude Code directly in your terminal
- Access to advanced Research
- Connect any content or tool through integrations
- Early access to advanced Claude features
- Priority access at high traffic times

Anthropic Claude  
\$0 (Free) / \$17<sup>2</sup> (Plus) / \$100 (Max)  
per Month

Give your team their time back

Whether you're a small business or a large enterprise, Perplexity Enterprise Pro enables your team to think strategically rather than search endlessly.

Get started

perplexity Free

perplexity Pro \$20/month

perplexity ENTERPRISE Pro \$40/month per seat or \$4000/year (quote 2024)

Get started for free

Get started

Get started

Perplexity  
\$0 (Free) / \$20 (Pro)  
per Month

Note: Excludes enterprise plans. 1. Grok pricing is bundled with X premium subscriptions. X premium subscriptions include additional benefits beyond improvements to Grok usage limits. 2. With annual discount. Source: OpenAI, X, Google, Anthropic, Perplexity websites (5/25)



注意：不包括企业计划。1. Grok 定价与 X premium 订阅捆绑在一起。X premium 订阅除了改进 Grok 使用限制外，还包括其他权益。2. 采用年度折扣。资料来源：OpenAI、X、Google、Anthropic、Perplexity 网站 (5/25)

# ...AI Monetization...Foundation Models = Developer API Fees Driving Monetization

## OpenAI ChatGPT, xAI Grok, Google Gemini, Anthropic Claude & Perplexity Developer API Pricing – 5/25, per Companies

GPT-4.1	GPT-4.1 mini	GPT-4.1 nano
Smartest model for complex tasks	Affordable model balancing speed and intelligence	Fastest, most cost effective model for low latency tasks
Price	Price	Price
Input: \$2.00 / M tokens	Input: \$5.40 / M tokens	Input: \$2.00 / M tokens
Context read: \$2.00 / M tokens	Context read: \$5.40 / M tokens	Context read: \$2.00 / M tokens
Output: \$5.00 / M tokens	Output: \$5.40 / M tokens	Output: \$2.00 / M tokens
Reasoning models for complex, multi-step problems		
OpenAI o3	OpenAI o4-mini	
Our most powerful reasoning model with leading performance on coding, math, science, and vision	Our faster, cost efficient reasoning model delivering strong performance on math, coding and vision	
Price	Price	
Input: \$2.00 / M tokens	Input: \$2.00 / M tokens	
Context read: \$2.00 / M tokens	Context read: \$2.00 / M tokens	
Output: \$5.00 / M tokens	Output: \$2.00 / M tokens	

**OpenAI ChatGPT**  
From \$0.40 (GPT-4.1 nano) to \$40 (o3) per 1MM Output Tokens

Models and Pricing	
An overview of our models' capabilities and their associated pricing. Our Grok 3 models come in two variants: a fast and a standard version. <a href="#">Details here.</a>	
	
<b>Grok 3</b> Our flagship model that excels at enterprise use cases like data extraction, coding, and text summarization. Possesses deep domain knowledge in finance, healthcare, law, and science.	<b>Grok 3 Mini (with Thinking)</b> A lightweight model that thinks before responding. Fast, smart, and great for logic-based tasks that do not require deep domain knowledge. The raw thinking traces are accessible.
<div>Example Usage</div>	<div>Example Usage</div>
<b>grok-3-beta</b>	<b>grok-3-mini-beta</b>
Input: \$1.00 / M	Input: \$1.00 / M
Output: \$1.00 / M	Output: \$1.00 / M

**xAI Grok**  
\$0.50 (grok-3-mini-beta) to \$25 (grok-3-fast) per 1MM Output Tokens

Get the best of Google AI, including Gemini Advanced and 2 TB of storage	
Gemini Advanced is free for students through Peak 2025. <a href="#">To be eligible, see here.</a>	
WITH A BUSINESS ACCOUNT	WITH A PREMIUM PLAN*
✓ 10 TB of storage	✓ 1 TB of storage
✓ Help with writing, planning, learning and more with Gemini	✓ Help with writing, planning, learning and more with Gemini
✓ Access to your personalized AI-powered research and writing assistant	✓ Gemini Advanced with access to our most capable models
	✓ Access to integration across Google apps with Gemini in Gmail, Docs, and more
	✓ Priority access to future innovations
	✓ Integrated AI that works with higher usage limits and premium features
	✓ Custom high-quality images and text. Get more when generation models in Gemini Advanced and other
	✓ And more premium benefits across Google
<div>Example Usage</div>	<div>Example Usage</div>
<b>gemini-1.5-flash-8b</b>	<b>gemini-1.5-pro</b>
Input: \$0.05 / M	Input: \$0.05 / M
Output: \$0.05 / M	Output: \$0.05 / M



**Google Gemini**  
\$0.15 (1.5 Flash-8B) to \$15 (2.5 Pro Preview) per 1MM Output Tokens<sup>1</sup>

# ...AI Monetization...Foundation Models = 开发者 API 费用推动货币化

## OpenAI ChatGPT, xAI Grok, Google Gemini, Anthropic Claude & Perplexity 开发者 API 定价 – 5/25, 每家公司

GPT-4.1	GPT-4.1 mini	GPT-4.1 nano
Smartest model for complex tasks	Affordable model balancing speed and intelligence	Fastest, most cost effective model for low latency tasks
Price	Price	Price
Input: \$2.00 / M tokens	Input: \$5.40 / M tokens	Input: \$2.00 / M tokens
Context read: \$2.00 / M tokens	Context read: \$5.40 / M tokens	Context read: \$2.00 / M tokens
Output: \$5.00 / M tokens	Output: \$5.40 / M tokens	Output: \$2.00 / M tokens
Reasoning models for complex, multi-step problems		
OpenAI o3	OpenAI o4-mini	
Our most powerful reasoning model with leading performance on coding, math, science, and vision	Our faster, cost efficient reasoning model delivering strong performance on math, coding and vision	
Price	Price	
Input: \$2.00 / M tokens	Input: \$2.00 / M tokens	
Context read: \$2.00 / M tokens	Context read: \$2.00 / M tokens	
Output: \$5.00 / M tokens	Output: \$2.00 / M tokens	

**OpenAI ChatGPT**  
从 0.40 美元（GPT-4.1 nano）到 40 美元（o3） / 每百万个输出 Token

Models and Pricing	
An overview of our models' capabilities and their associated pricing. Our Grok 3 models come in two variants: a fast and a standard version. <a href="#">Details here.</a>	
	
<b>Grok 3</b> Our flagship model that excels at enterprise use cases like data extraction, coding, and text summarization. Possesses deep domain knowledge in finance, healthcare, law, and science.	<b>Grok 3 Mini (with Thinking)</b> A lightweight model that thinks before responding. Fast, smart, and great for logic-based tasks that do not require deep domain knowledge. The raw thinking traces are accessible.
<div>Example Usage</div>	<div>Example Usage</div>
<b>grok-3-beta</b>	<b>grok-3-mini-beta</b>
Input: \$1.00 / M	Input: \$1.00 / M
Output: \$1.00 / M	Output: \$1.00 / M

**xAI Grok**  
\$0.50 (grok-3-mini-beta) to \$25 (grok-3-fast) per 1MM 输出 Tokens

Get the best of Google AI, including Gemini Advanced and 2 TB of storage	
Gemini Advanced is free for students through Peak 2025. <a href="#">To be eligible, see here.</a>	
WITH A BUSINESS ACCOUNT	WITH A PREMIUM PLAN*
✓ 10 TB of storage	✓ 1 TB of storage
✓ Help with writing, planning, learning and more with Gemini	✓ Help with writing, planning, learning and more with Gemini
✓ Access to your personalized AI-powered research and writing assistant	✓ Gemini Advanced with access to our most capable models
	✓ Access to integration across Google apps with Gemini in Gmail, Docs, and more
	✓ Priority access to future innovations
	✓ Integrated AI that works with higher usage limits and premium features
	✓ Custom high-quality images and text. Get more when generation models in Gemini Advanced and other
	✓ And more premium benefits across Google
<div>Example Usage</div>	<div>Example Usage</div>
<b>gemini-1.5-flash-8b</b>	<b>gemini-1.5-pro</b>
Input: \$0.05 / M	Input: \$0.05 / M
Output: \$0.05 / M	Output: \$0.05 / M

**Google Gemini**  
\$0.15 (1.5 Flash-8B) 到 \$15 (2.5 Pro Preview) per 1MM Output Tokens<sup>1</sup>

Pricing			
Non-Reasoning Models			
Models optimized for fast, cost effective search and information retrieval			
Model	Price	Context Window	Price per 1000 Tokens
Claude 3.7 Sonnet	\$3 / MTok	200K	\$0.30
Claude 3.5 Haiku	\$0.80 / MTok	200K	\$0.08
Claude 3 Opus	\$15 / MTok	200K	\$1.50
Reasoning Models			
Model	Price	Context Window	Price per 1000 Tokens
Claude 3.7 Sonnet	\$3 / MTok	200K	\$0.30
Claude 3.5 Haiku	\$0.80 / MTok	200K	\$0.08
Claude 3 Opus	\$15 / MTok	200K	\$1.50

**Anthropic Claude**  
From \$1.25 (Claude 3 Haiku) to \$75 (Claude 3 Opus) per 1MM Output Tokens

Pricing			
Non-Reasoning Models			
Models optimized for fast, cost effective search and information retrieval			
Model	Price	Context Window	Price per 1000 Tokens
Gemini 1.5 Flash-8B	\$0.05 / M	1M	\$0.05
Gemini 1.5 Pro	\$0.05 / M	1M	\$0.05
Gemini 2.0 Flash-8B	\$0.05 / M	1M	\$0.05
Gemini 2.0 Pro	\$0.05 / M	1M	\$0.05
Reasoning Models			
Model	Price	Context Window	Price per 1000 Tokens
Gemini 1.5 Flash-8B	\$0.05 / M	1M	\$0.05
Gemini 1.5 Pro	\$0.05 / M	1M	\$0.05
Gemini 2.0 Flash-8B	\$0.05 / M	1M	\$0.05
Gemini 2.0 Pro	\$0.05 / M	1M	\$0.05

**Perplexity**  
\$1 (Sonar) to \$15 (Sonar Pro) per 1MM Output Tokens

Pricing			
Non-Reasoning Models			
Models optimized for fast, cost effective search and information retrieval			
Model	Price	Context Window	Price per 1000 Tokens
Claude 3.7 Sonnet	\$3 / MTok	200K	\$0.30
Claude 3.5 Haiku	\$0.80 / MTok	200K	\$0.08
Claude 3 Opus	\$15 / MTok	200K	\$1.50
Reasoning Models			
Model	Price	Context Window	Price per 1000 Tokens
Claude 3.7 Sonnet	\$3 / MTok	200K	\$0.30
Claude 3.5 Haiku	\$0.80 / MTok	200K	\$0.08
Claude 3 Opus	\$15 / MTok	200K	\$1.50

**Anthropic Claude**  
From \$1.25 (Claude 3 Haiku) 到 \$75 ( Claude 3 Opus) 每 100 万个输出 Token

Pricing			
Non-Reasoning Models			
Models optimized for fast, cost effective search and information retrieval			
Model	Price	Context Window	Price per 1000 Tokens
Gemini 1.5 Flash-8B	\$0.05 / M	1M	\$0.05
Gemini 1.5 Pro	\$0.05 / M	1M	\$0.05
Gemini 2.0 Flash-8B	\$0.05 / M	1M	\$0.05
Gemini 2.0 Pro	\$0.05 / M	1M	\$0.05
Reasoning Models			
Model	Price	Context Window	Price per 1000 Tokens
Gemini 1.5 Flash-8B	\$0.05 / M	1M	\$0.05
Gemini 1.5 Pro	\$0.05 / M	1M	\$0.05
Gemini 2.0 Flash-8B	\$0.05 / M	1M	\$0.05
Gemini 2.0 Pro	\$0.05 / M	1M	\$0.05

**困惑度**  
\$1 (Sonar) 到 \$15 (Sonar Pro) 每 100 万个输出 Token

1. Gemini prices by prompt size. Gemini 1.5 Flash-8B = \$0.15 per 1MM tokens for prompts ≤128K tokens; Gemini 2.5 Pro Preview = \$15 per 1MM tokens for prompts >200K tokens. Source: OpenAI, X, Google, Anthropic, Perplexity websites (5/25)

1. Gemini 价格取决于提示大小。 Gemini 1.5 Flash-8B = 每个提示 1MM tokens 的价格为 \$0.15≤128K tokens ; Gemini 2.5 Pro Preview = \$ 每个提示 1MM tokens 的价格为 15 美元 >200K tokens。来源：OpenAI、X、Google、Anthropic、Perplexity 网站 (5/25)

*AI – New Entrants =  
Rapid Revenue Growth*

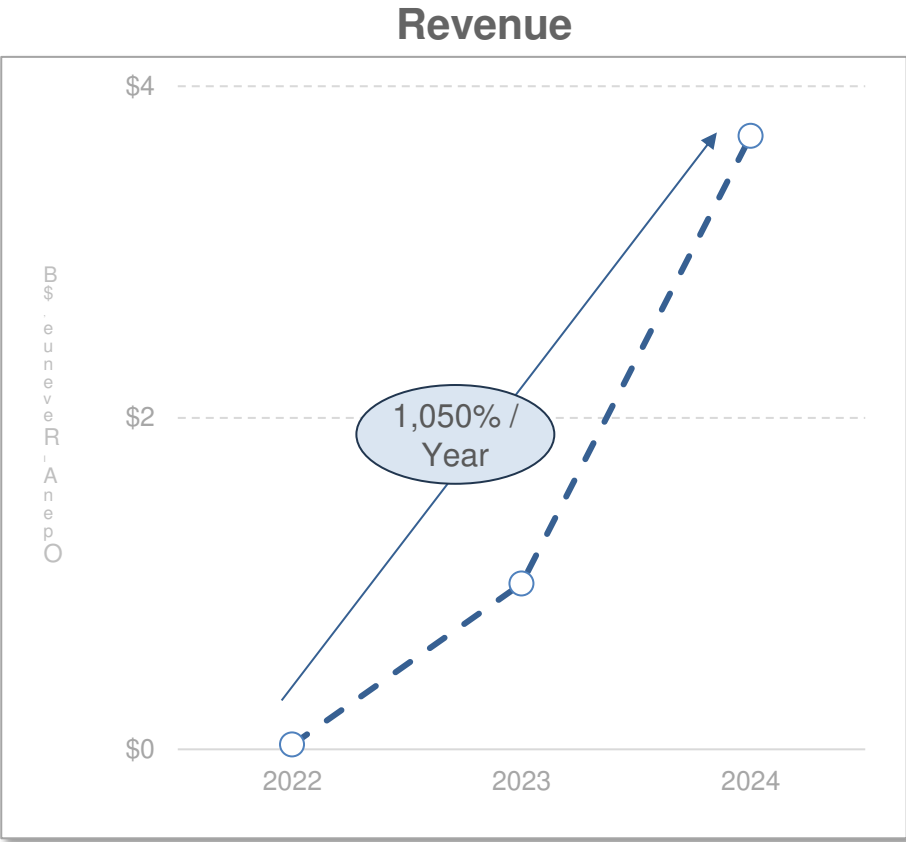
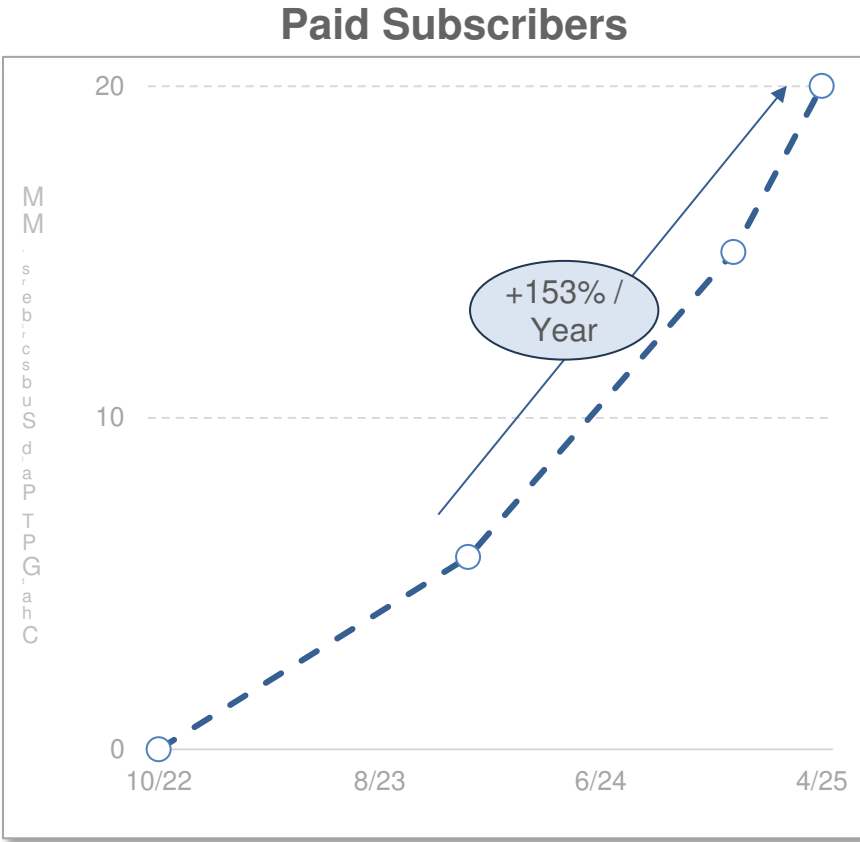
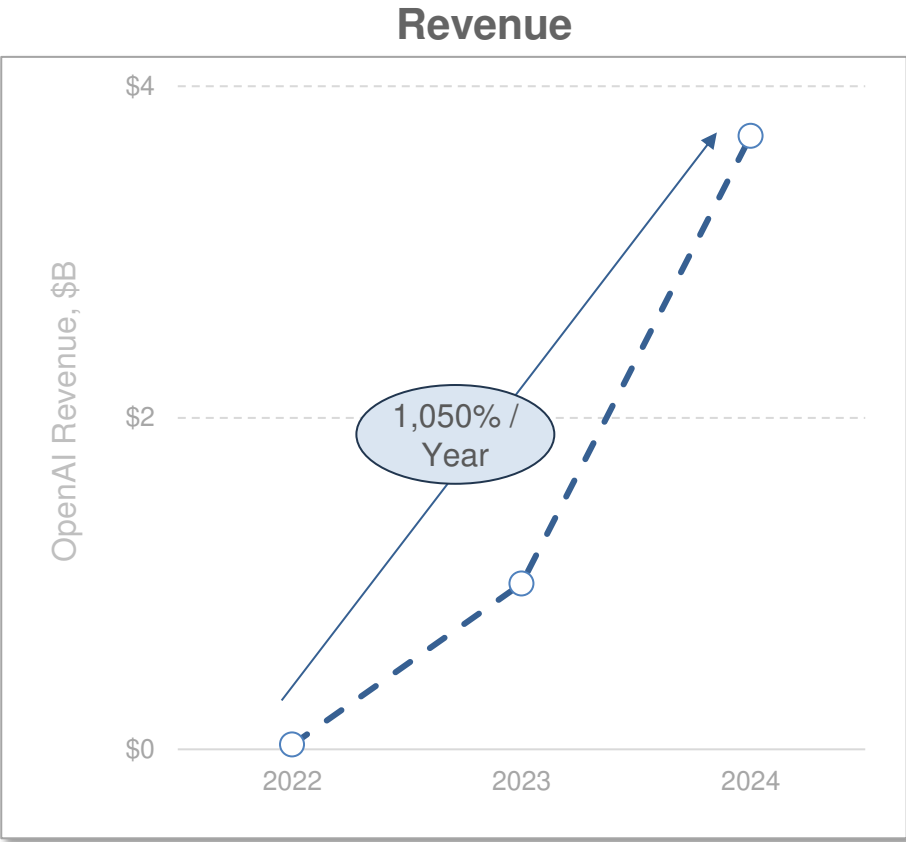
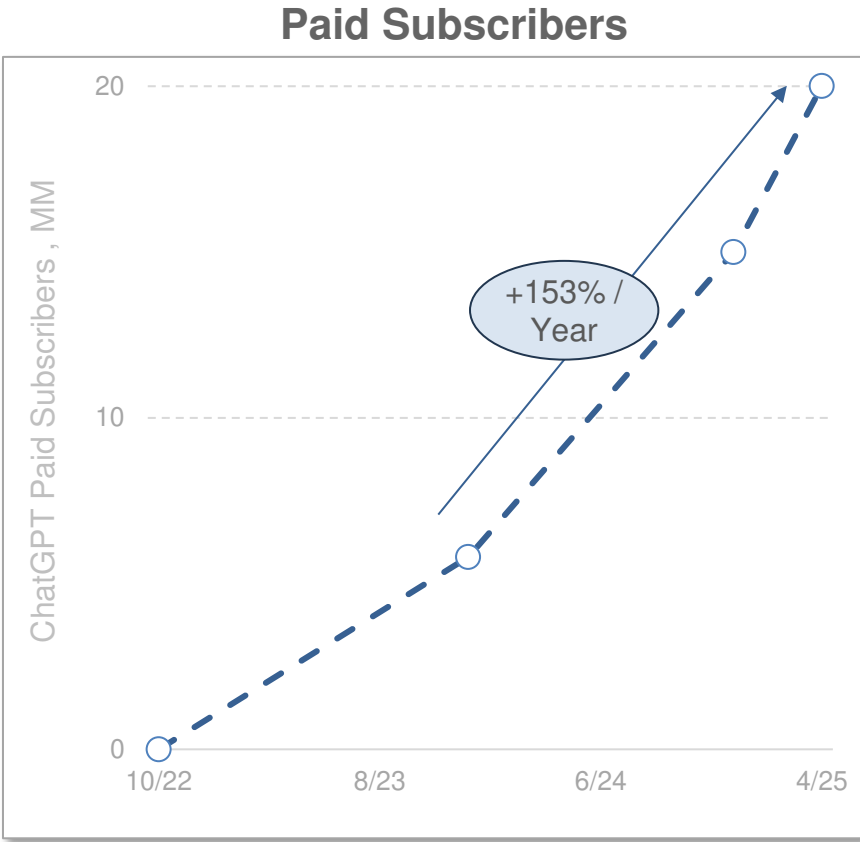
AI – 新进入者 =  
收入快速增长

AI Monetization – Foundation Models =  
OpenAI Revenue +1,050% Annually to \$3.7B

AI 货币化 – 基础模型 =OpenAI 收入 +1，年增长  
率为 50%，达到 37 亿美元

ChatGPT Paid Subscribers (MM) & Revenue (\$B) – 10/22-4/25,  
per OpenAI & *The Information*

ChatGPT 付费订阅者（百万） & 收入（十亿美元） – 10/22-4/25，数据  
来源：OpenAI & The Information

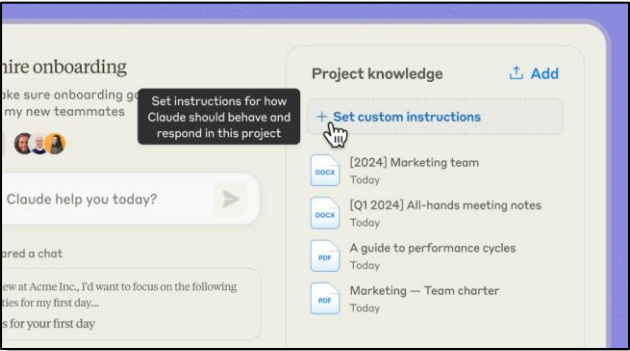


Source: OpenAI disclosures (as of 4/25), The Information (4/25) ([link](#), [link](#), [link](#) & [link](#))

Source: OpenAI disclosures (as of 4/25), The Information (4/25) ([link](#), [link](#), [link](#) & [link](#))

# AI Monetization – API & Generative Search = Anthropic Annualized Revenue +20x to \$2B in Eighteen Months

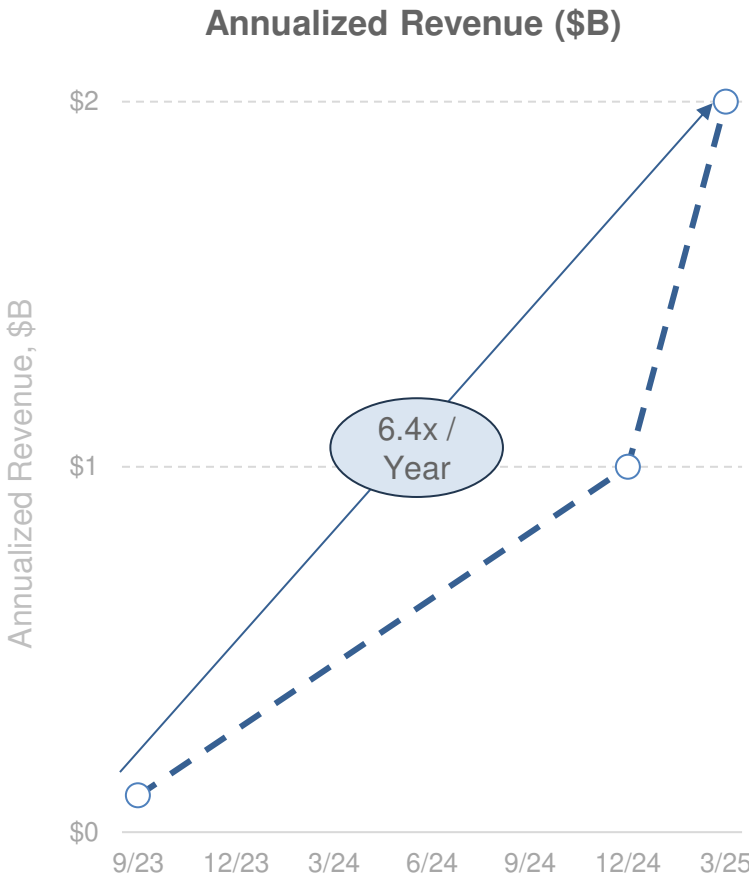
## Anthropic: API & Generative Search – 9/23-3/25, per Reuters, Bloomberg & CNBC



We've developed Claude 3.7 Sonnet with a different philosophy from other reasoning models on the market. Just as humans use a single brain for both quick responses and deep reflection, we believe reasoning should be an integrated capability of frontier models rather than a separate model entirely. This unified approach also creates a more seamless experience for users...

...we've optimized somewhat less for math and computer science competition problems, and instead shifted focus towards real-world tasks that better reflect how businesses actually use LLMs.

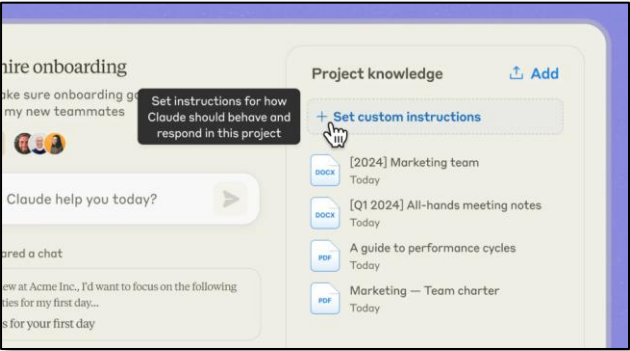
- Anthropic Press Release, 2/25



Source: Anthropic; Reuters, 'Anthropic forecasts more than \$850 mln in annualized revenue rate by 2024-end – report' (12/23) ([link](#)); Bloomberg, 'Anthropic Finalizes Megaround at \$61.5 Billion Valuation' (3/25) ([link](#)); CNBC, 'Anthropic closes \$2.5 billion credit facility as Wall Street continues plunging money into AI boom' (5/25) ([link](#))

# AI 货币化 – API 和生成式搜索 =Anthropic 年化收入 +20x 在 18 个月内达到 20 亿美元

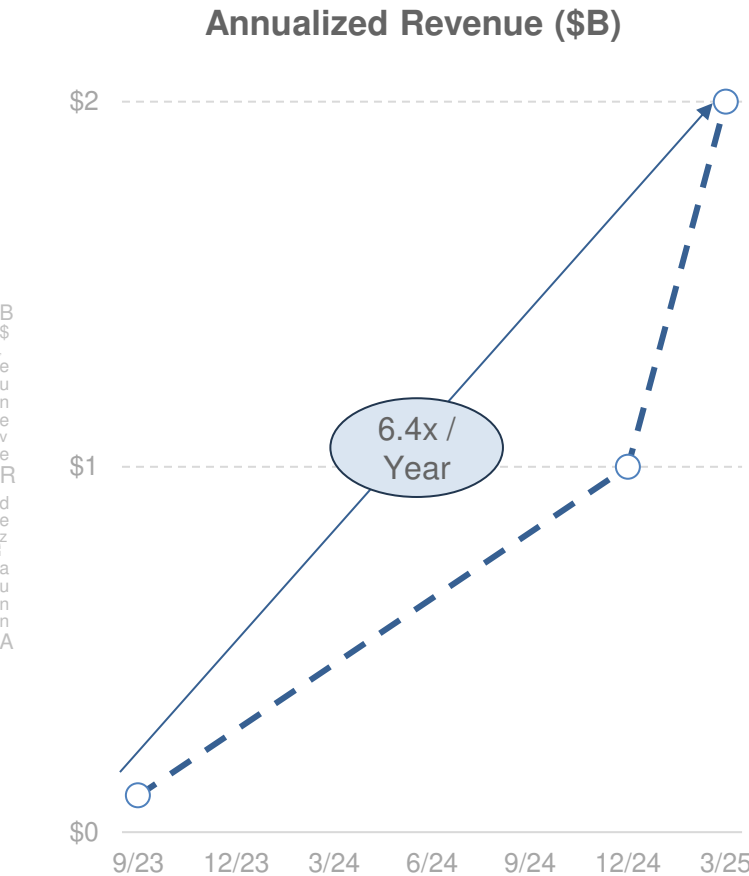
## Anthropic: API 和生成式搜索 – 9/23-3/25, 据路透社、彭博社和 CNBC 报道



我们开发的 Claude 3.7 Sonnet 采用了与市场上其他推理模型不同的理念。正如人类使用一个大脑进行快速反应和深度思考一样，我们认为推理应该是前沿模型的集成能力，而不是完全独立的模型。这种统一的方法也为用户创造了更无缝的体验 ……

… 我们对数学和计算机科学竞赛问题的优化有所减少，而是将重点转向更好地反映企业实际使用法学硕士的现实世界任务。

- Anthropic 新闻稿, 2/25



Source: Anthropic; Reuters, 'Anthropic forecasts more than \$850 mln in annualized revenue rate by 2024-end – report' (12/23) ([link](#)); Bloomberg, 'Anthropic Finalizes Megaround at \$61.5 Billion Valuation' (3/25) ([link](#)); CNBC, 'Anthropic closes \$2.5 billion credit facility as Wall Street continues plunging money into AI boom' (5/25) ([link](#))

# AI Monetization – Generative Search = Perplexity Annualized Revenue +7.6x to \$120MM in Fourteen Months

# AI 货币化 – 生成式搜索 =Perplexity 年度化收入 +7.6 倍，在 14 个月内达到 1.2 亿美元

## Perplexity: Generative Search – 3/24-5/25, per Perplexity & Bloomberg

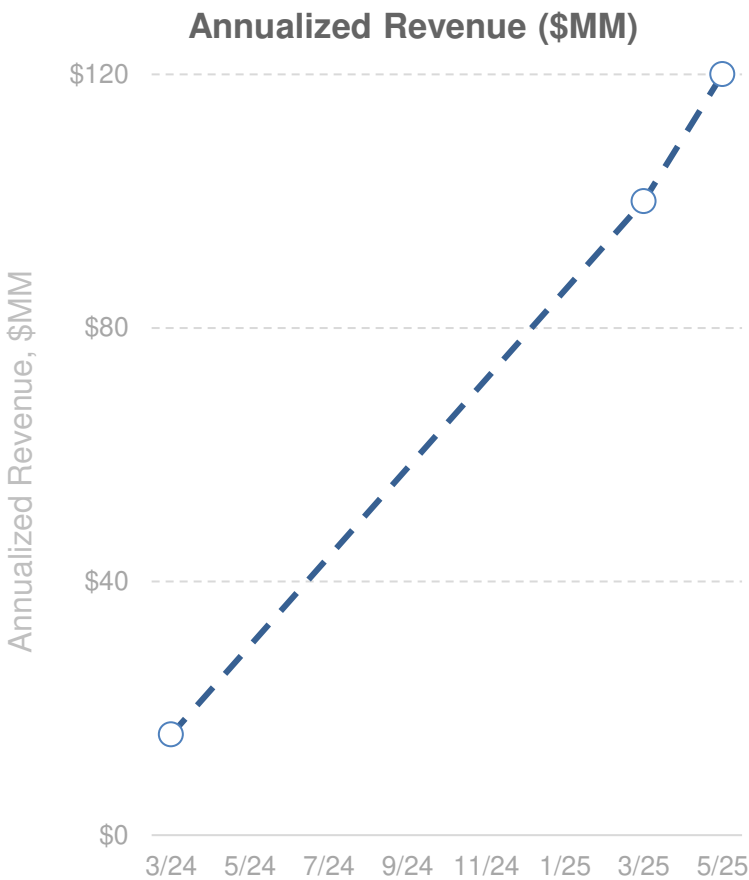


*Perplexity is best described as an answer engine. You ask it a question, you get an answer. Except the difference is, all the answers are backed by sources. This is like how an academic writes a paper...What makes humans special is that we are creatures of curiosity. We need to expand on that and discover more knowledge using the power of AI.*

- Perplexity Co-Founder & CEO Aravind Srinivas, 6/24

What if accessing information felt like talking to a personal research assistant?

- Perplexity Co-Founder & CEO Aravind Srinivas, 5/25



Note: 3/24 annualized revenue figure is an estimate per Perplexity Co-Founder & CEO Aravind Srinivas's 3/25 LinkedIn post saying 'Perplexity has crossed \$100m in annualized revenue...6.3x growth Y/Y and remains highly under monetized.'  
Source: Lex Fridman Podcast (6/24), UC Berkeley (5/25), LinkedIn (3/25), Bloomberg, 'AI Startup Perplexity Nears Funding at \$14 Billion Value' (5/25) (link)

## Perplexity: 生成式搜索 – 3/24-5/25，根据 Perplexity 和彭博社

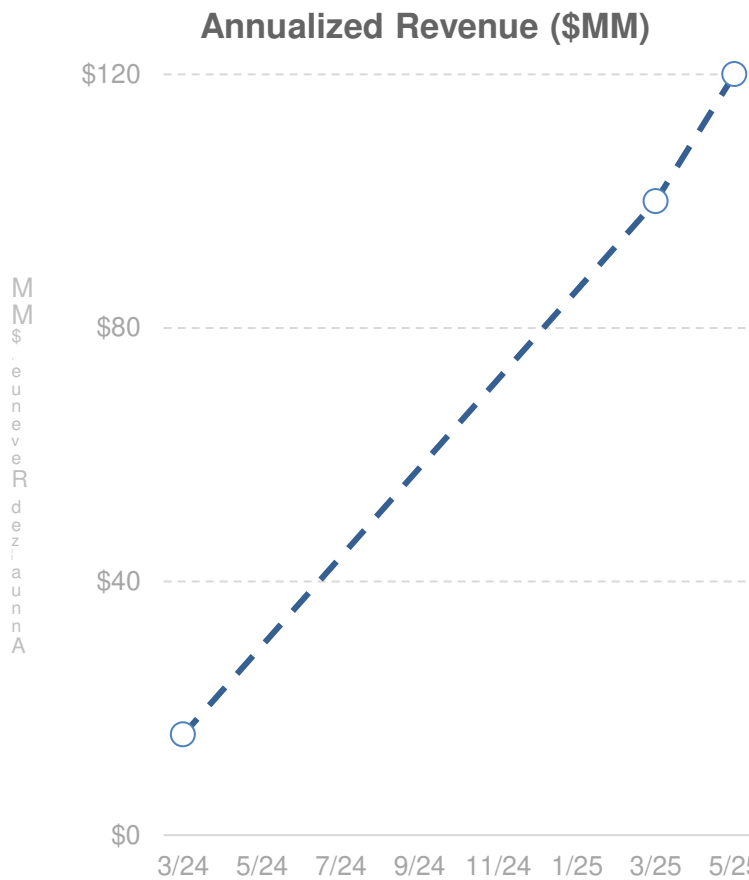


Perplexity 最好的描述是答案引擎。你问它一个问题，你会得到一个答案。不同之处在于，所有答案都有来源支持。这就像学者写论文一样 …… 人类的特别之处在于我们是充满好奇心的生物。我们需要扩展这一点，并利用人工智能的力量发现更多知识。

- Perplexity Co-Founder & CEO Aravind Srinivas, 6/24

如果获取信息感觉像和私人研究助理交谈一样呢？

- Perplexity 联合创始人兼 CEO Aravind Srinivas, 5/25



注：3/24 年度化收入数据是根据 Perplexity 联合创始人兼首席执行官 Aravind Srinivas 3/25 的 LinkedIn 帖子估算的，该帖子称 “Perplexity 的年度化收入已超过 1 亿美元 …… 同比增长 6.3 倍，但货币化程度仍然很低。” 来源：Lex Fridman Podcast (6/24)、加州大学伯克利分校 (5/25)、LinkedIn (3/25)、彭博社，《人工智能初创公司 Perplexity 接近以 \$14 0 亿美元估值融资》(5/25) (link)

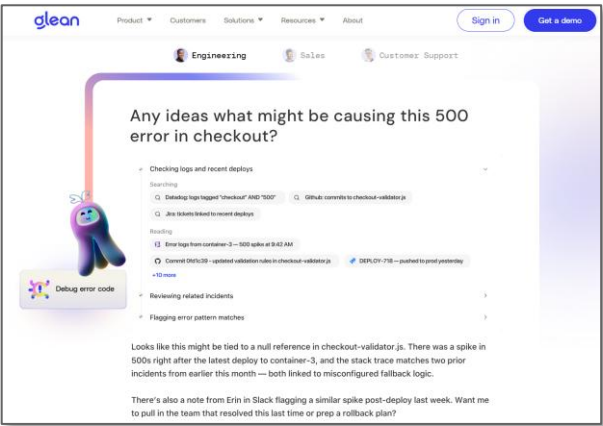


# AI Monetization – Enterprise Search + Agents = Glean Annualized Revenue +10x to \$100MM in Twenty-Four Months

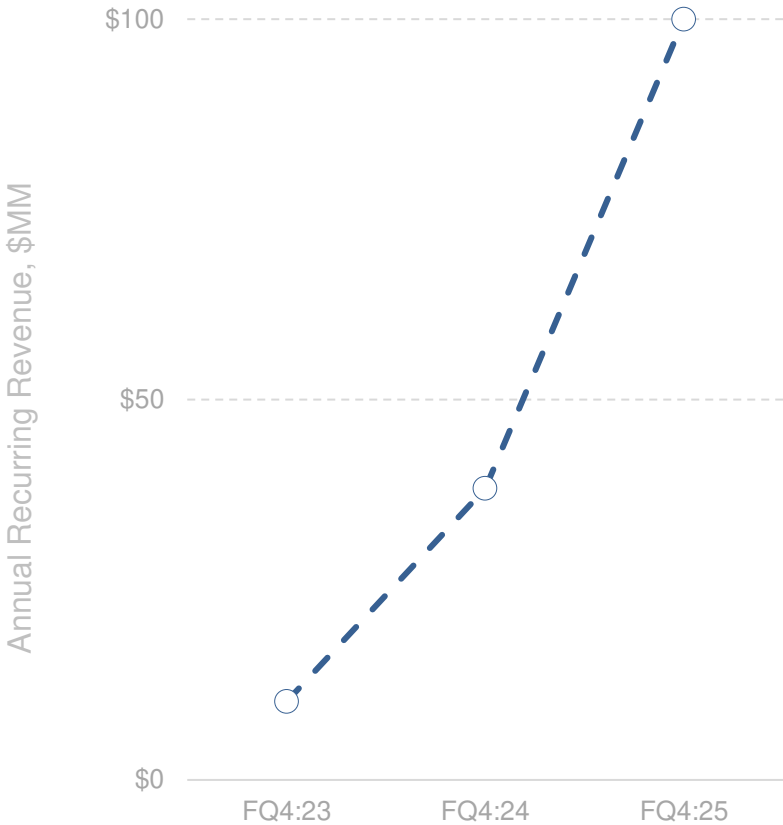
# AI 货币化 – 企业搜索 + 代理 =Glean 年度化收入 +10x 在 24 个月内达到 1 亿美元

Glean – FQ4:23-FQ4:25, per Glean

Glean – FQ4:23-FQ4:25, 根据 Glean



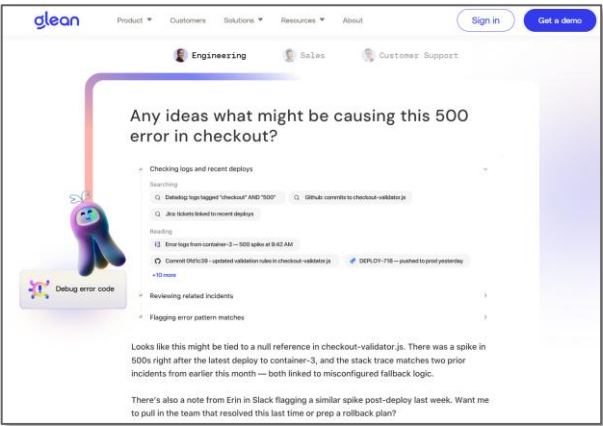
Annual Recurring Revenue (ARR) (\$MM)



*We’re honored to help some of the world’s largest companies adopt AI to transform their businesses. To truly unlock new levels of creativity, productivity, and operational efficiency, AI needs to draw on the full picture of an organization’s knowledge – and it needs to be accessible by everyone.*

*You shouldn’t have to be a prompt engineering expert to find answers, generate content, and automate work with AI.*

**- Glean Co-Founder & CEO Arvind Jain (9/24)**

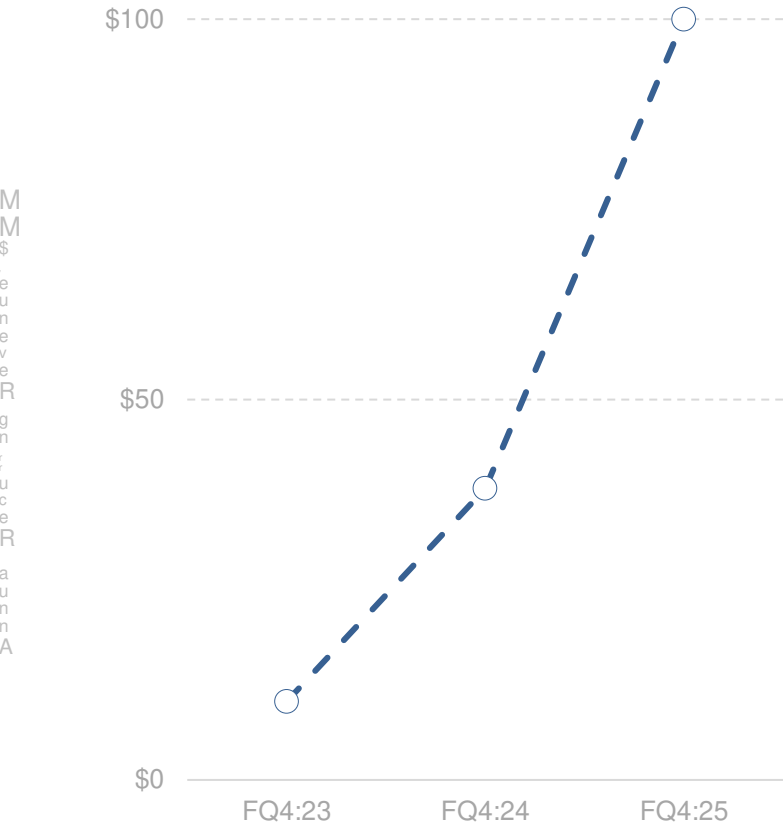


我们很荣幸能够帮助一些世界上最大的公司采用 AI 来转型他们的业务。为了真正释放创造力、生产力和运营效率的新水平，AI 需要利用组织知识的全貌 – 并且需要每个人都可以访问。

您不应该非得成为提示工程专家才能使用 AI 查找答案、生成内容和自动化工作。

- Glean 联合创始人兼首席执行官 Arvind Jain (9/24)

年度经常性收入 (ARR) ( 百万美元 )



Note: Glean’s fiscal year ends in January. Source: Glean (2/25, 11/24)

Note: Glean’s fiscal year ends in January. Source: Glean (2/25, 11/24)

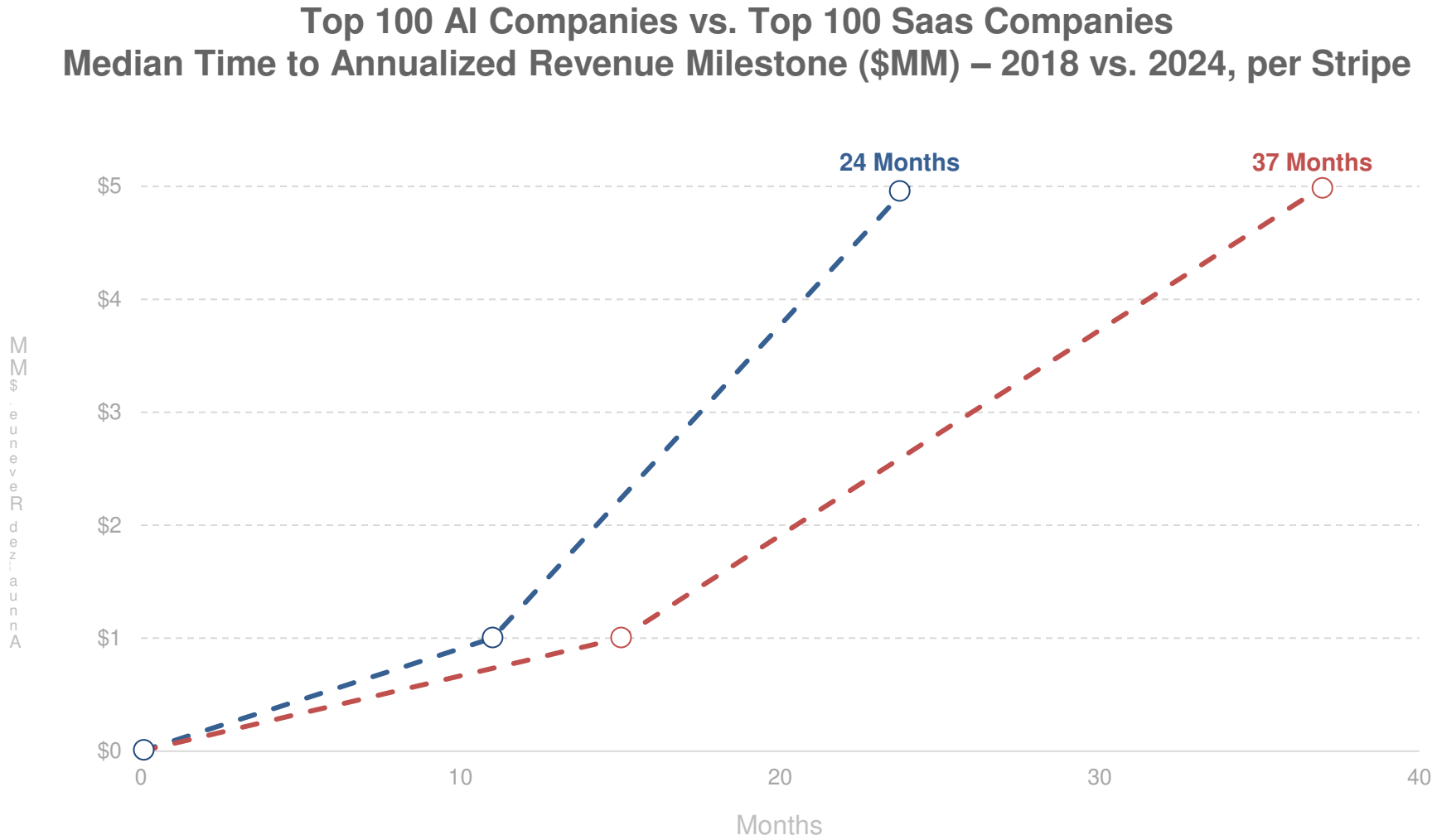


AI Monetization – 2024 vs. 2018 =  
35% Faster Ramp to \$5MM ARR vs. SaaS Comparables, per Stripe

AI 货币化 – 2024 vs. 2018 = 比 SaaS 同类产品更快 35% 达到 500 万美元  
ARR, 根据 Stripe 数据



Source: Stripe Annual Letter (2/25)



Source: Stripe Annual Letter (2/25)

AI Monetization Possibilities =

New Entrants & / Or **Tech Incumbents**?

AI 货币化可能性 =

新的进入者和 / 或科技巨头?

*AI – Tech Incumbents =*

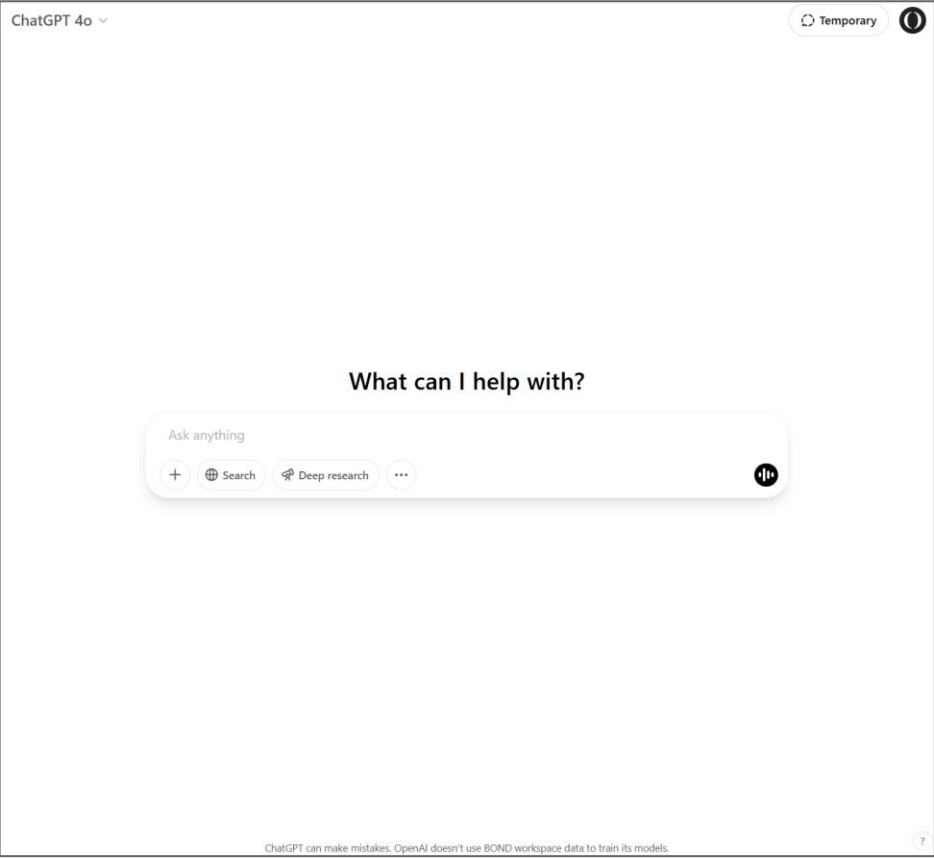
*Broad & Steady Product / Feature Rollouts*

AI – 科技巨头 =

广泛且稳定的产品 / 功能发布

# Tech Incumbents = Optimizing Product Distribution to Roll Out AI

While ChatGPT Has 800MM+ Users Via Its Website & App...



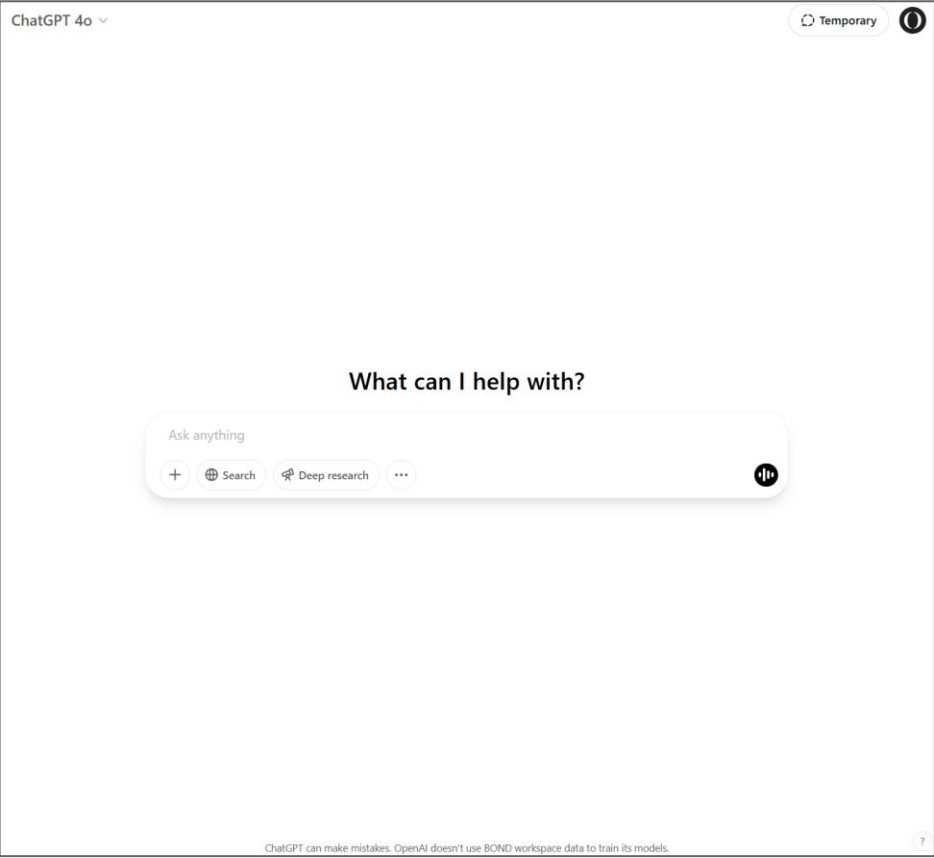
...Tech Incumbents Have Billions of Global Users on Devices & Platforms With Ongoing AI Product Rollouts

	Google 4.9B Search Users, 3B+ Android Users, 1.5B AI Overviews Users & 1B+ Assistant Devices
	Meta Users* 3.4B+
	Apple Devices** 2.35B
	TikTok Users*** 1B+
	Microsoft 1B LinkedIn Members & 400MM+ Office 365 Paid Seats
	Spotify Users 678MM
	Amazon 600MM+ Alexa Devices & 200MM+ Prime Subscribers
	X Users 500MM+
	Canva Users 230MM+

*\*Meta includes Facebook, Instagram, WhatsApp, & Messenger. \*\*Apple includes iPhones, iPads, Macs, & other Apple devices worldwide. \*\*\*As of 2021; no more recent company data available. Note: Some figures are estimates based off past company disclosures & web traffic / purchase history analytics. Different companies may define 'users' differently based on frequency. Source: Statcounter (2/25), Google (5/25), Meta 10Q (4/25), Apple (1/25), TikTok (7/21), LinkedIn (5/25), Microsoft (1/24), Spotify (5/25), Amazon (2/25 & 10/24), Elon Musk via X (7/23), Canva (4/25), OpenAI disclosures (4/25), Wikimedia Commons*

# Tech Incumbents = Optimizing Product Distribution to Roll Out AI

While ChatGPT Has 800MM+ Users Via Its Website & App...



...Tech Incumbents Have Billions of Global 拥有数十亿用户的设备和平台上的科技巨头正在持续推出人工智能产品

	Google 4.9B 搜索用户 , AI 概览用户	3B+ Android Users, 1.5B & 1B+ Assistant Devices
	Meta 用户 *	3.4B+
	Apple 设备 **	2.35B
	TikTok 用户 ***	1B+
	Microsoft 1B LinkedIn 会员 rs &	400MM+ Office 365 Paid Seats
	Spotify 用户	678MM
	Amazon 600MM+ Alexa 设备 es &	200MM+ Prime 订阅 cribers
	X 用户	500MM+
	Canva 用户	230MM+

*\*Meta 包括 Facebook 、 Instagram 、 WhatsApp 和 Messenger。 \*\*Apple 包括 iPhone 、 iPad 、 Mac 和全球其他 Apple 设备。 \*\*\* 截至 2021 年 ; 没有更新的公司数据可用。 注意 : 有些数据是根据过去的公司披露和网络流量 / 购买历史分析估算的。 Different companies may define 'users' differently based on 频率。 来源 : Statcounter (2/25) 、 Google (5/25) 、 Meta 10Q (4/25) 、 Apple (1/25) 、 TikTok (7/21) 、 LinkedIn (5/25) 、 Microsoft (1/24) 、 Spotify (5/25) 、 Amazon (2/25 & 10/24) 、 Elon Musk via X (7/23) 、 Canva (4/25) 、 OpenAI disclosures (4/25) 、 Wikimedia Commons*

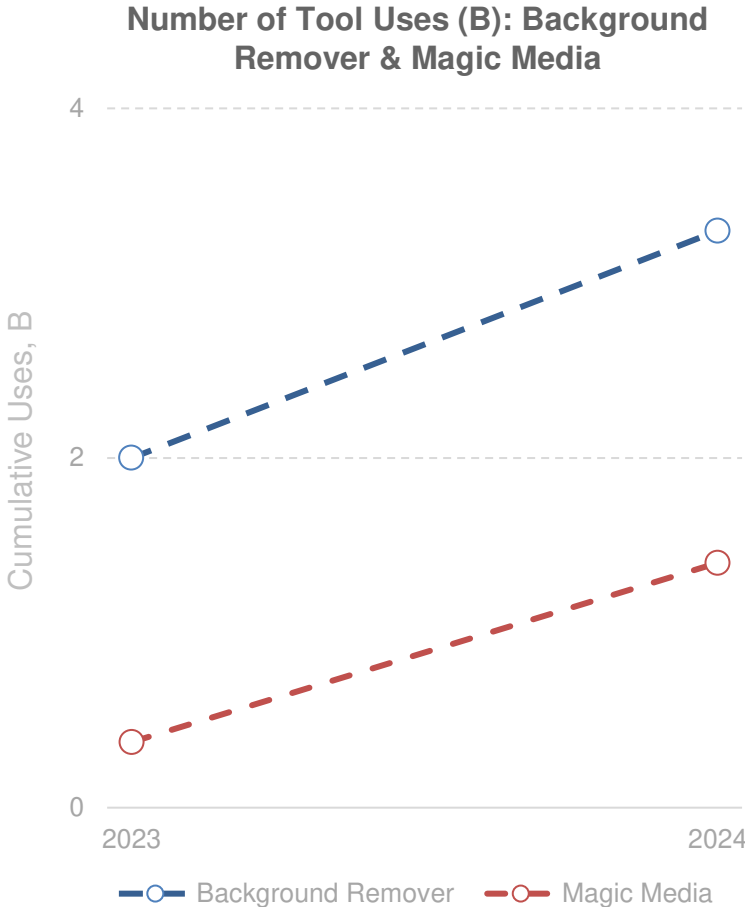
Tech Incumbent AI Rollouts =  
Canva – Background Remover & Magic Media (12/19)

Canva Background Remover & Magic Media – 2023-2024, per Canva



One of our community’s favorite Canva features has been the one-click image Background Remover, launched in December 2019...[to] wild success and community love.  
- Canva Press Release, 9/22

Magic Media lets you turn your imagination into reality by watching your words transform into stunning, one-of-a-kind images – and now videos and graphics, too...In less than a year since launching Magic Media’s text to image, we’ve been overwhelmed by our community’s enthusiastic response, with almost 290 million images being created and applied to a range of practical use cases from social media posts to presentations, business flyers, and even logos.  
- Canva Press Release, 10/24



Source: Canva announcements & press releases (2022-2024)

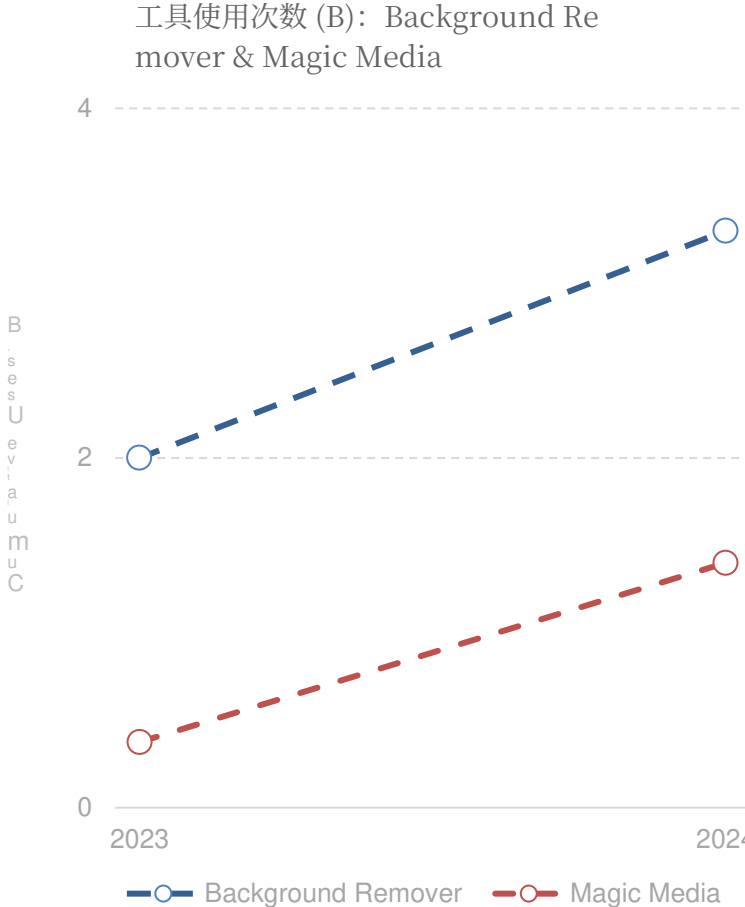
科技巨头人工智能发布 =Canva – 背景移除器和魔法媒体  
（ 12/19 ）

Canva 背景移除器和魔法媒体 – 2023-2024， 据 Canva



我们社区最受欢迎的 Canva 功能之一是一键图像背景移除器，于 2019 年 12 月推出 ...[到 ] 获得了巨大的成功和社区的喜爱。 - Canva 新闻稿， 9/22

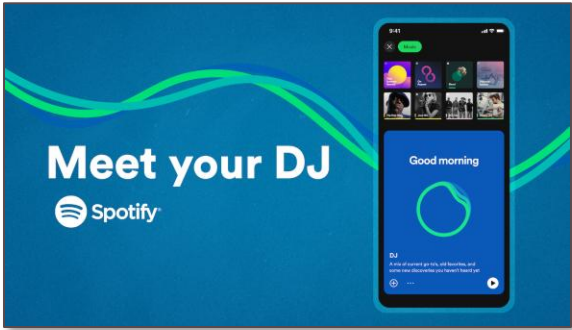
通过观看您的文字转化为令人惊叹的、独一无二的图像，Magic Media 让您将想象力变为现实 —— 现在还有视频和图形 …… 自从推出 Magic Media 的文本到图像功能以来不到一年，我们对社区的热烈反响感到不知所措，已经创建了近 2.9 亿张图像，并应用于从社交媒体帖子到演示文稿、商业传单甚至徽标等各种实际用例。 - Canva 新闻稿， 10/24



Source: Canva announcements & press releases (2022-2024)

Tech Incumbent AI Rollouts = Spotify – AI DJ (2/23)

Spotify AI DJ – 2/23-5/25, per Spotify



AI DJ and music videos...are truly moving averages...  
AI DJ, we're seeing amazing results, not just on quantitative metrics, but also on quality metrics, how people feel about Spotify, what they say they love about Spotify.

- Spotify Co-Founder & CEO Daniel Ek, 11/24

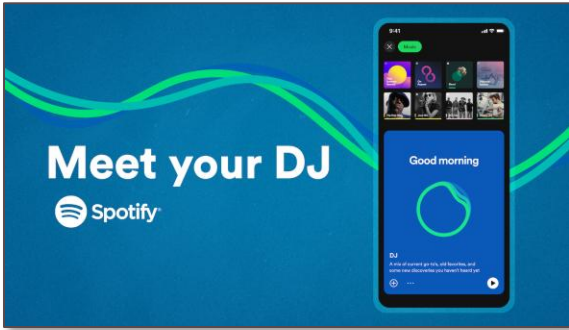
Back in 2018, we said something internally that still holds true today: machine learning – what most people called AI back then – was the product...  
AI is really the next step in evolution, where machine learning allows personalization, AI also allows for real time interactivity and reasoning on top of your data.

- Spotify Co-President, Chief Product & Technology Officer Gustav Söderström, 4/25

Source: Company announcements (2/23, 5/23, 8/23, 11/24, 4,25, 5/25)

科技巨头人工智能发布 = Spotify – AI DJ (2/23)

Spotify AI DJ – 2/23-5/25, 根据 Spotify



AI DJ 和音乐视频 …… 确实是移动平均线 …… AI DJ, 我们看到了惊人的结果, 不仅在定量指标上, 而且在质量指标上, 人们对 Spotify 的感觉, 他们说他们喜欢 Spotify 的什么。

- Spotify 联合创始人兼 CEO Daniel Ek, 11/24

早在 2018 年, 我们在内部说过一些至今仍然有效的話: 机器学习 – 当时大多数人称之为 AI – 是产品 …… AI 实际上是进化的下一步, 机器学习可以实现个性化, AI 还可以根据您的数据进行实时交互和推理。

- Spotify 联合总裁、首席产品和技术官 Gustav Söderström, 4/25

来源: 公司公告 (2/23, 5/23, 8/23, 11/24, 4,25, 5/25)

Microsoft: Copilot – 8/23-12/24, per Microsoft

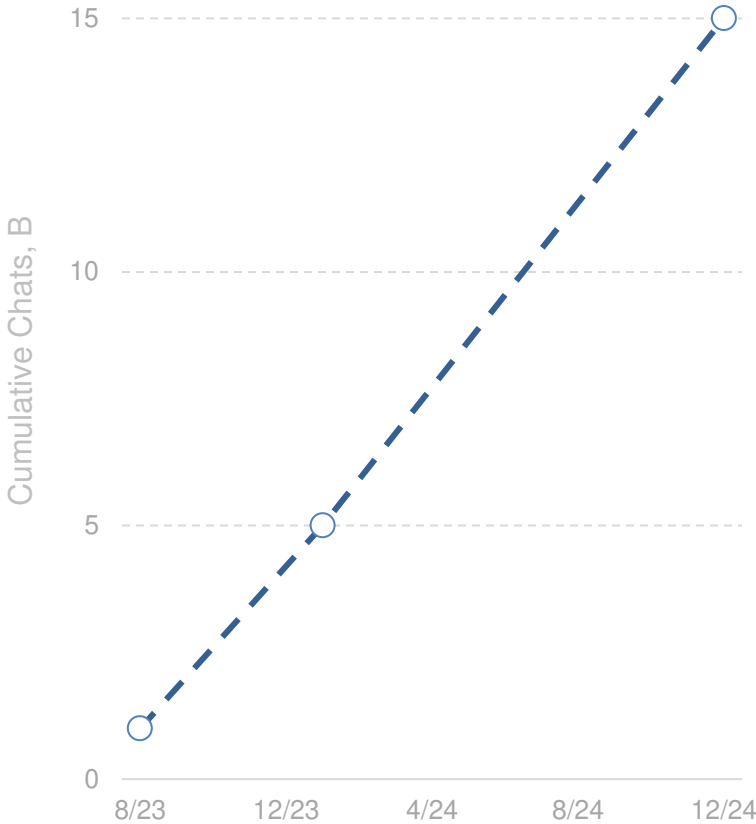


*To empower people to unlock the joy of discovery, feel the wonder of creation and better harness the world’s knowledge, today we’re improving how the world benefits from the web by reinventing the tools billions of people use every day, the search engine and the browser.*

*Today, we’re launching an all new, AI-powered Bing search engine and Edge browser, available in preview now at Bing.com, to deliver better search, more complete answers, a new chat experience and the ability to generate content. We think of these tools as an AI copilot for the web.*

**- Official Microsoft Blog, 2/23**

Microsoft Copilot Cumulative Chats Held (B)



Note: We assume zero users in the launch month. We assume 15B cumulative chats as of 12/24 due to Microsoft’s 1/24 announcement of 5B cumulative chats, and 12/24 announcement of 10B more chats being held in 2024. We assume the Verge’s announcement of ‘There have also been over 1 billion chats on Bing Chat’ as of 8/23 is wholly inclusive of Copilot chat volumes as of that date. Source: Microsoft announcements & earnings reports, The Verge citing Microsoft disclosures (8/23)

Microsoft: Copilot – 8/23-12/24, 根据微软数据

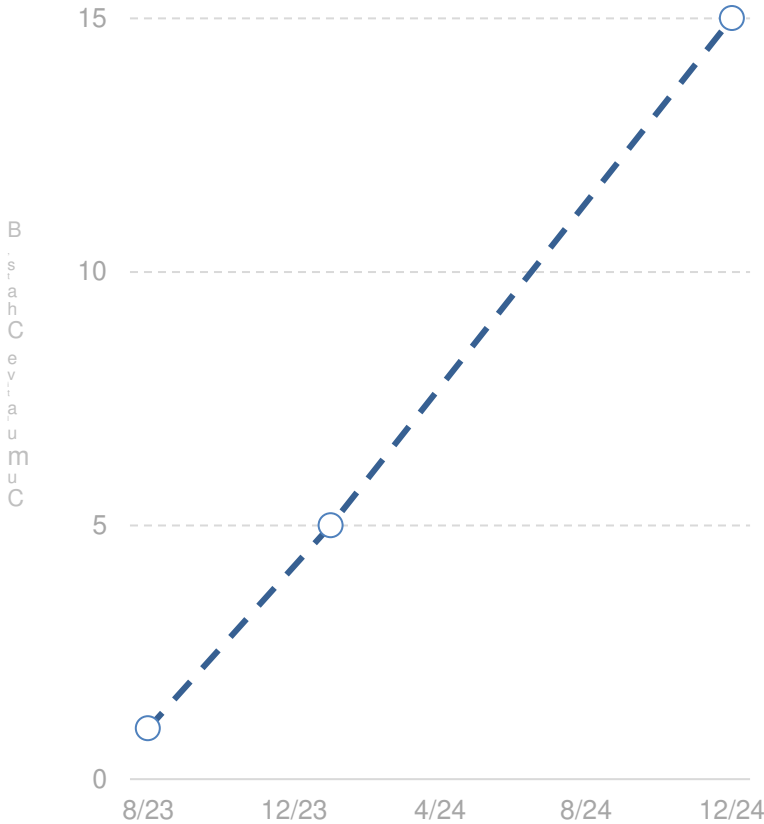


为了让人们能够享受发现的乐趣，感受创造的奇迹，更好地利用世界的知识，今天我们正在改进世界从网络中受益的方式，通过重塑数十亿人每天使用的工具、搜索引擎和浏览器。

今天，我们推出了一个全新的、由人工智能驱动的 Bing 搜索引擎和 Edge 浏览器，现在可以在 Bing.com 上预览，以提供更好的搜索、更完整的答案、全新的聊天体验以及生成内容的能力。我们将这些工具视为网络的 AI 副驾驶。

**- Official Microsoft Blog, 2/23**

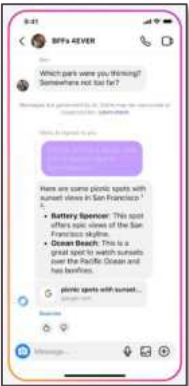
Microsoft Copilot Cumulative Chats Held (B)



注意：我们假设发布月份的用户数为零。由于微软在 1 月 24 日宣布累计聊天次数为 50 亿次，以及在 12 月 24 日宣布 2024 年将再进行 100 亿次聊天，因此我们假设截至 12 月 24 日累计聊天次数为 150 亿次。我们假设 The Verge 在 8 月 23 日宣布的“Bing Chat 上的聊天次数已超过 10 亿次”完全包含截至该日期的 Copilot 聊天量。来源：微软公告和收益报告，The Verge 引用微软披露的信息 (8/23)。



Meta Platforms: Meta AI – 11/23-4/25, per Meta Platforms

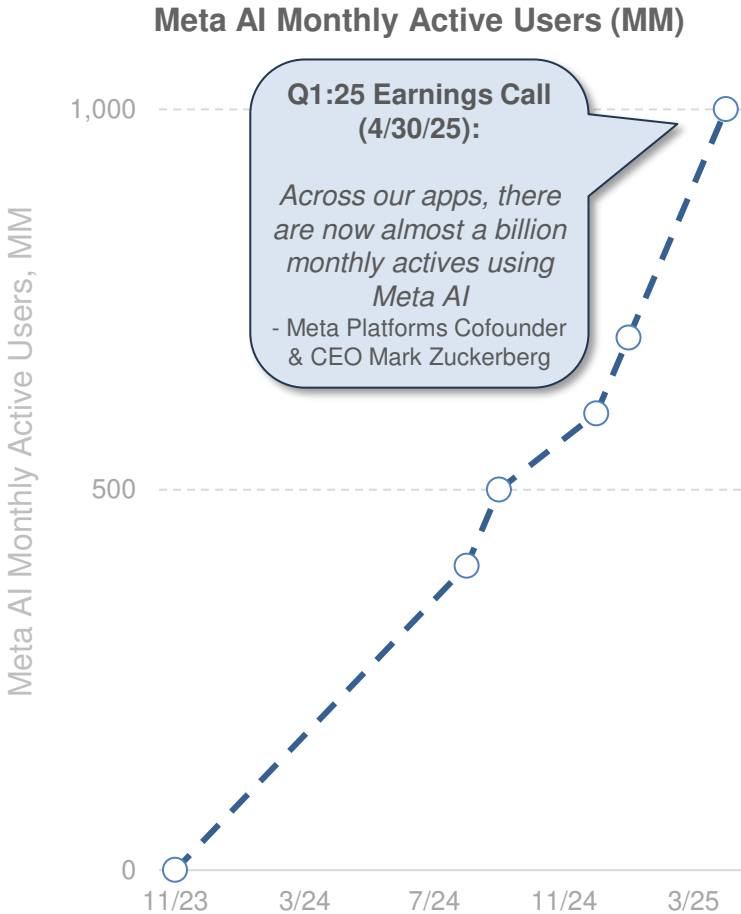


*I expect that this is going to be the year when a highly intelligent and personalized AI assistant reaches more than 1 billion people, and I expect Meta AI to be that leading AI assistant. Meta AI is already used by more people than any other assistant...*

*...I also expect that 2025 will be the year when it becomes possible to build an AI engineering agent that has coding and problem-solving abilities of around a good mid-level engineer...*

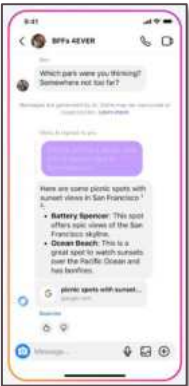
*...Whichever company builds [a high-skill AI engineering agent] first, I think it's going to have a meaningful advantage in deploying it to advance their AI research and shape the field.*

**- Meta Platforms CEO Mark Zuckerberg, 1/25**



Note: We assume zero users in 11/23 per Meta’s 12/23 blog post noting, ‘To chat with our AIs, start a new message and select “Create an AI chat” on Instagram, Messenger or WhatsApp. They’re now available to anyone in the US.’ Source: Meta Platforms announcements & earnings reports

Meta Platforms: Meta AI – 11/23-4/25, 根据 Meta Platforms

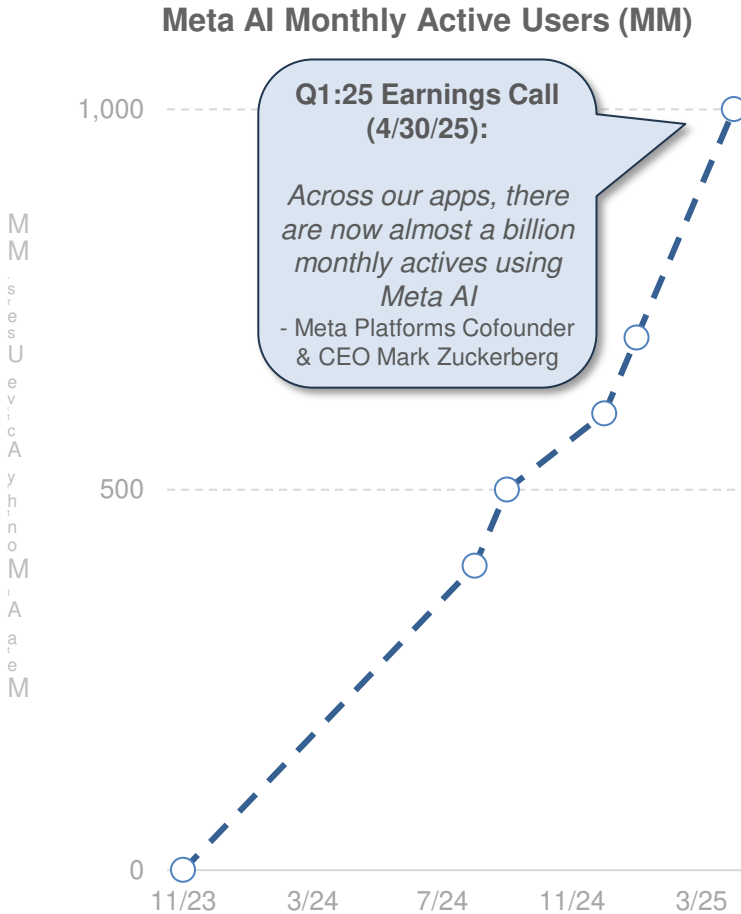


我预计今年将出现一个高度智能化和个性化的 AI 助手，其用户将超过 10 亿，并且我预计 Meta AI 将成为领先的 AI 助手。Meta AI 已经被比任何其他助手更多的人使用 ……*people than any other assistant...*

*...I also expect that 2025 will be the year when it becomes possible to build an AI engineering agent that has coding and problem-solving abilities of around a good mid-level engineer...*

*...Whichever company builds [a high-skill AI engineering agent] first, I think it's going to have a meaningful advantage in deploying it to advance their AI research and shape the field.*

**- Meta Platforms CEO Mark Zuckerberg, 1/25**



注意：根据 Meta 12/23 的博客文章，我们假设 11/23 的用户数为零，该文章指出，“要与我们的 AI 聊天，请开始一条新消息，然后在 Instagram、Messenger 或 WhatsApp 上选择“创建 AI 聊天”。它们现在可供美国任何人使用。”资料来源：Meta Platforms 公告和收益报告

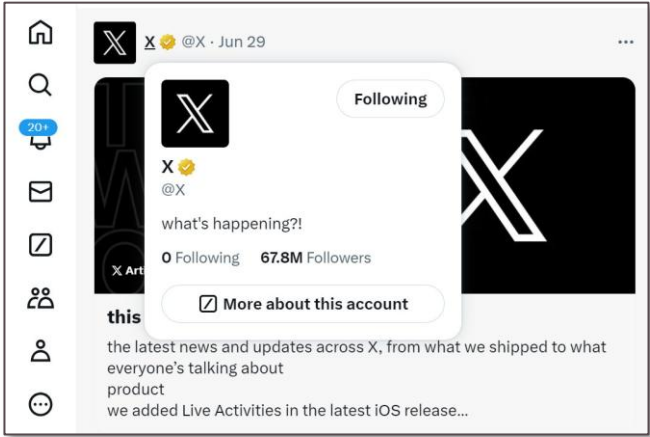


Tech Incumbent AI Rollouts =  
X – Grok (11/23)

科技巨头 AI 产品发布 =X –  
Grok (11/23)

X: Grok – 12/24-4/25, per xAI & Similarweb

X: Grok –12/24-4/25, 根据 xAI 和 Similarweb



*The mission of xAI and Grok is to understand the universe.  
We want to answer the biggest questions.*

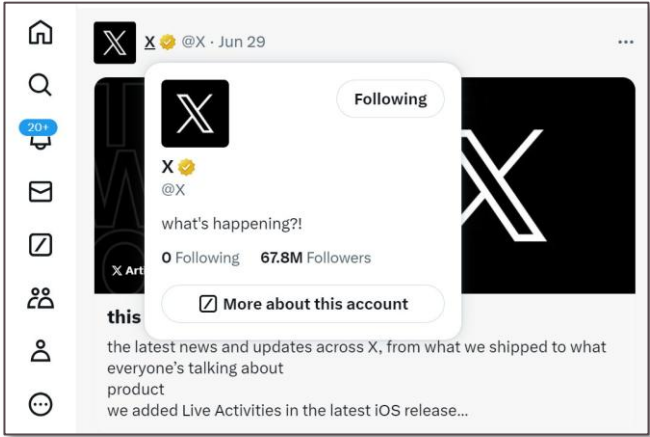
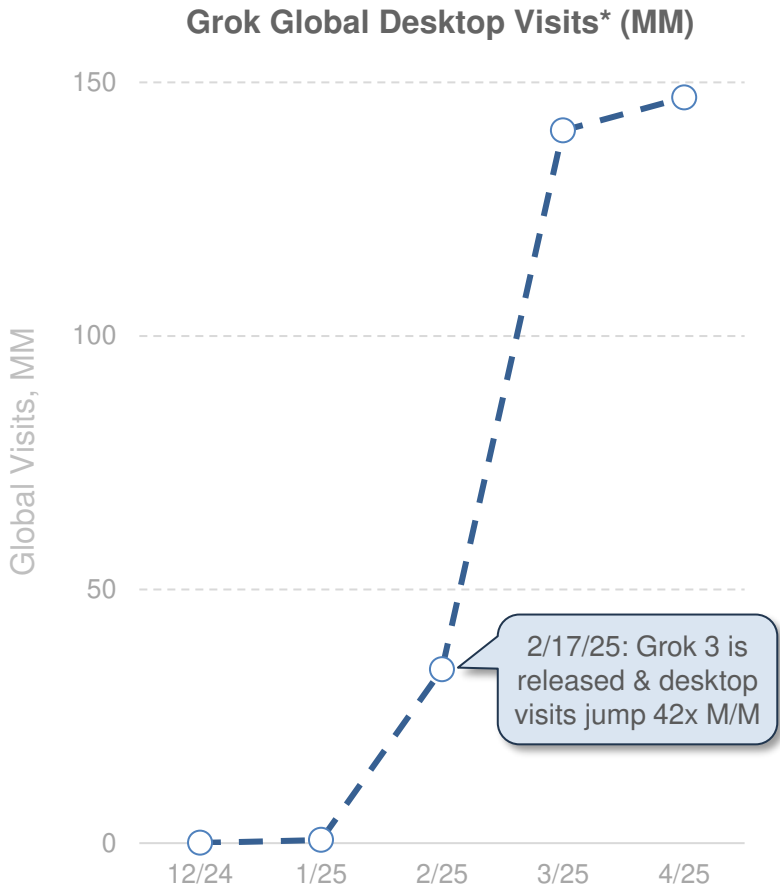
**- xAI Founder & CEO Elon Musk, 2/25**

*AI with Grok is getting very good...it's important that AI be  
programmed with good values, especially truth-seeking  
values. This is, I think, essential for AI safety...*

*...Remember these words: We must have a maximally truth-  
seeking AI.*

**- xAI Founder & CEO Elon Musk, 5/25**

\*Excludes X visits. China data may be subject to informational limitations due to government restrictions. Source: xAI announcements & investor filings; Elon Musk; Fox News; Similarweb (5/25)



xAI 和 Grok 的使命是了解宇宙。我们想要回答最大的问题。

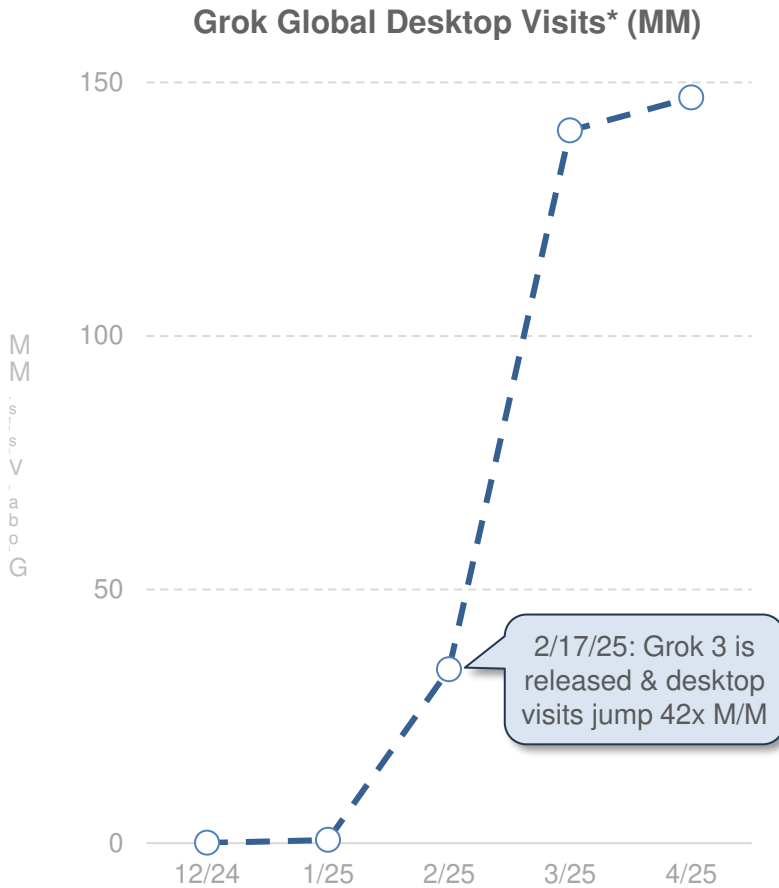
**- xAI 创始人兼 CEO Elon Musk, 2/25**

拥有 Grok 的 AI 变得非常好……重要的是 AI 要用良好的价值观  
进行编程，尤其是寻求真理的价值观。我认为，这对于 AI 安全至关  
重要……

…记住这些话：我们必须拥有一个最大限度地追求真理的 -  
seeking AI。

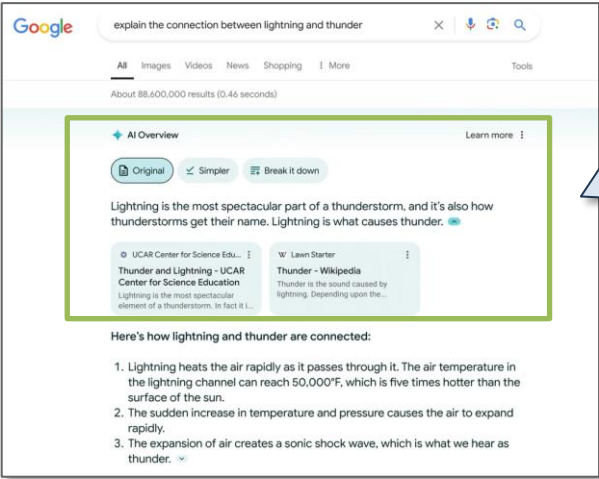
**- xAI 创始人兼首席执行官 Elon Musk, 5/25**

\* 不包括 X 的访问量。由于政府限制，中国的数据可能受到信息限制。来源：xAI 公告和投资者备案；Elon Musk；Fox News；Similarweb (5/25)



# Tech Incumbent AI Rollouts = Google – Gemini & AI Overviews (12/23)

## Alphabet: Gemini & AI Overviews – 3/25-5/25, per Alphabet & Business Insider

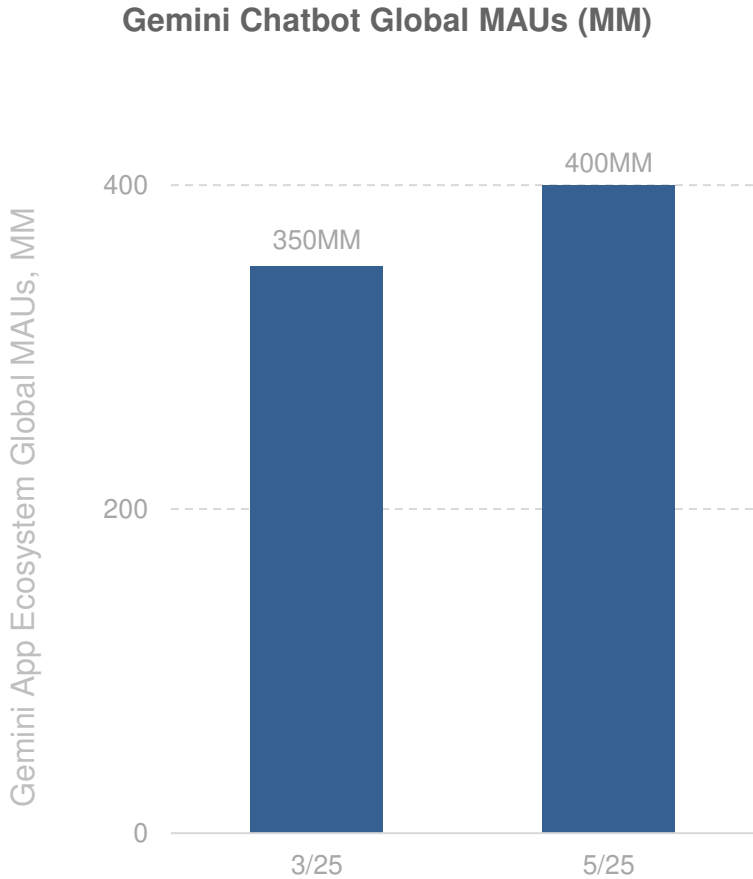


*Our differentiated, full stack approach to AI continues to be central to our growth. This quarter was super exciting as we rolled out Gemini 2.5, our most intelligent AI model, which is achieving breakthroughs in performance, and it’s widely recognized as the best model in the industry.*

**- Alphabet CEO Sundar Pichai, 4/25**

Google Gemini is a family of multimodal AI models, capable of understanding and generating various types of data including text, code, audio, images, and video.

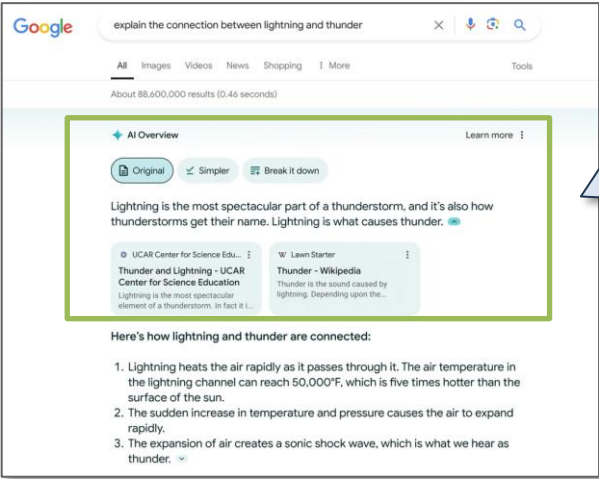
Source: Google Gemini



Note: Gemini launched 12/23...App launched 2/24. Data shown for apps in Gemini ecosystem. User counts may differ from those as measured by third-party data providers / panels like Similarweb & Sensor Tower as they measure only visits to desktop sites and standalone mobile apps, respectively. Source: Google announcements (4/25 & 5/25) & Business Insider, "Google's Gemini usage is skyrocketing, but rivals like ChatGPT and Meta AI are still blowing it out of the water" (4/25)

# Tech Incumbent AI Rollouts = Google – Gemini & AI Overviews (12/23)

## Alphabet: Gemini 和 AI Overviews – 3/25-5/25, 根据 Alphabet 和 Business Insider

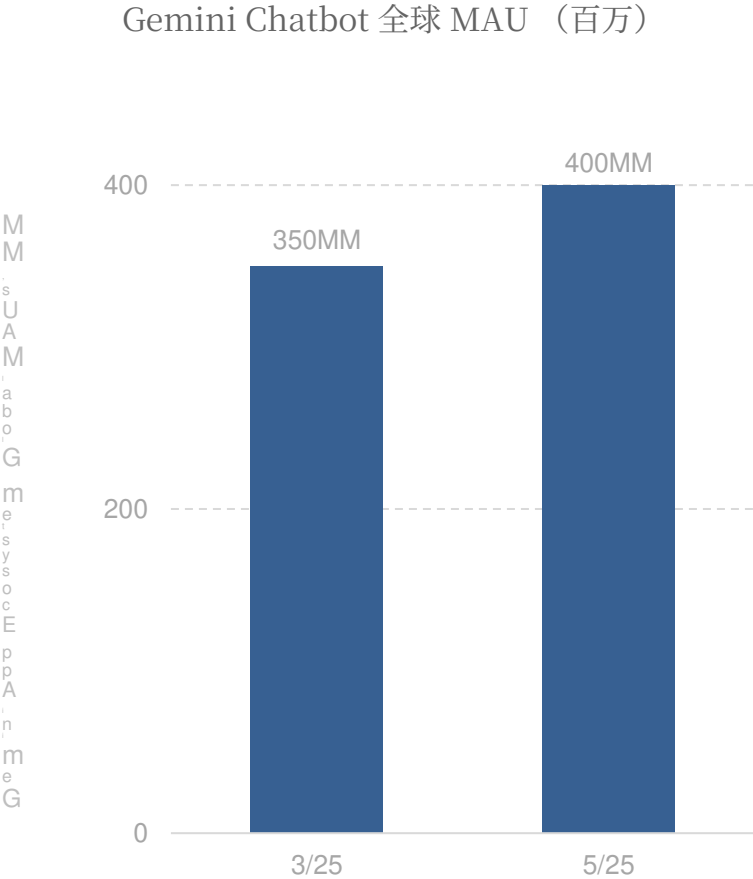


我们与众不同的、完整的 AI 方法仍然是我们增长的核心。本季度非常令人兴奋，因为我们推出了 Gemini 2.5，我们最智能的 AI 模型，它在性能上取得了突破，并且被广泛认为是业内最好的模型。

**- Alphabet CEO Sundar Pichai, 4/25**

Google Gemini 是一系列多模态 AI 模型，能够理解和生成各种类型的数据，包括文本、代码、音频、图像和视频。

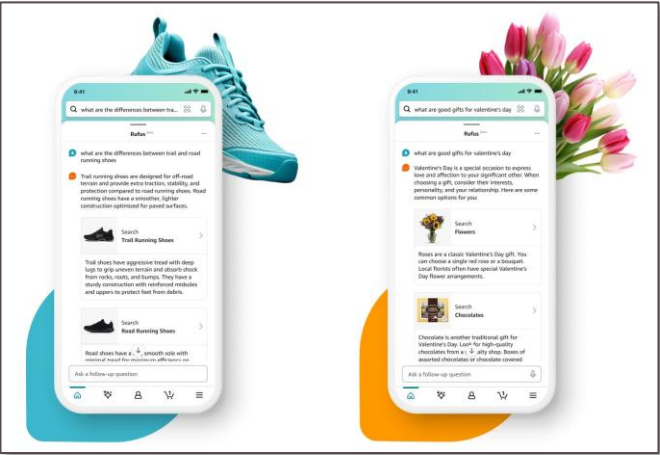
Source: Google Gemini



注意：Gemini 于 12/23 发布 …App 于 2/24 发布。数据显示的是 Gemini 生态系统中的应用。用户数量可能与 Similarweb 和 Sensor Tower 等第三方数据提供商 / 面板的测量结果不同，因为它们分别只测量桌面网站和独立移动应用的访问量。来源：Google 公告（4/25 和 5/25）和 Business Insider, “Google 的 Gemini 使用量正在飞速增长，但 ChatGPT 和 Meta AI 等竞争对手仍然遥遥领先”（4/25）

# Tech Incumbent AI Rollouts = Amazon – Rufus (2/24)

## Amazon: Rufus – 12/22-3/25, per Amazon & Morgan Stanley Estimates

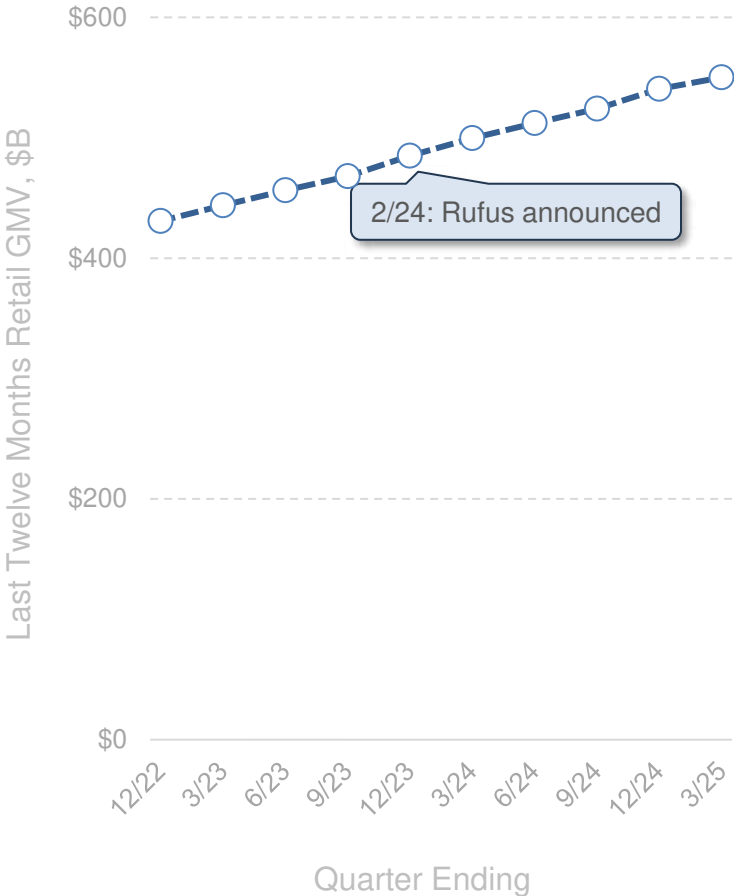


*We have so many customers now who just use Rufus to help them find a quick fact about a product. They also use Rufus to figure out how to summarize customer reviews, so they don't have to read 100 customer reviews to get a sense of what people think about that product...the personalization keeps getting much better...*

*...And so, we expect throughout 2025, that the number of occasions where you're not sure what you want to buy and you want help from Rufus are going to continue to increase and be more and more helpful to customers.*

- Amazon CEO Andy Jassy, 2/25

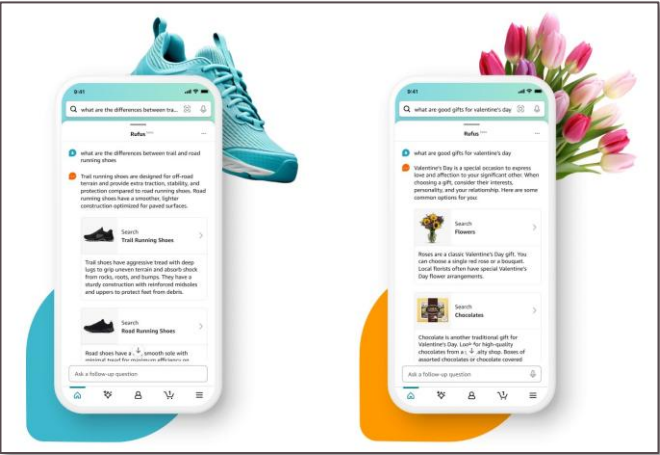
Amazon North America Retail Estimated Gross Merchandise Value (\$B), Last 12 Months



Source: Amazon; Morgan Stanley estimates

# 科技巨头人工智能发布 = Amazon – Rufus (2/24)

## Amazon: Rufus – 12/22-3/25, 根据亚马逊和摩根士丹利估计

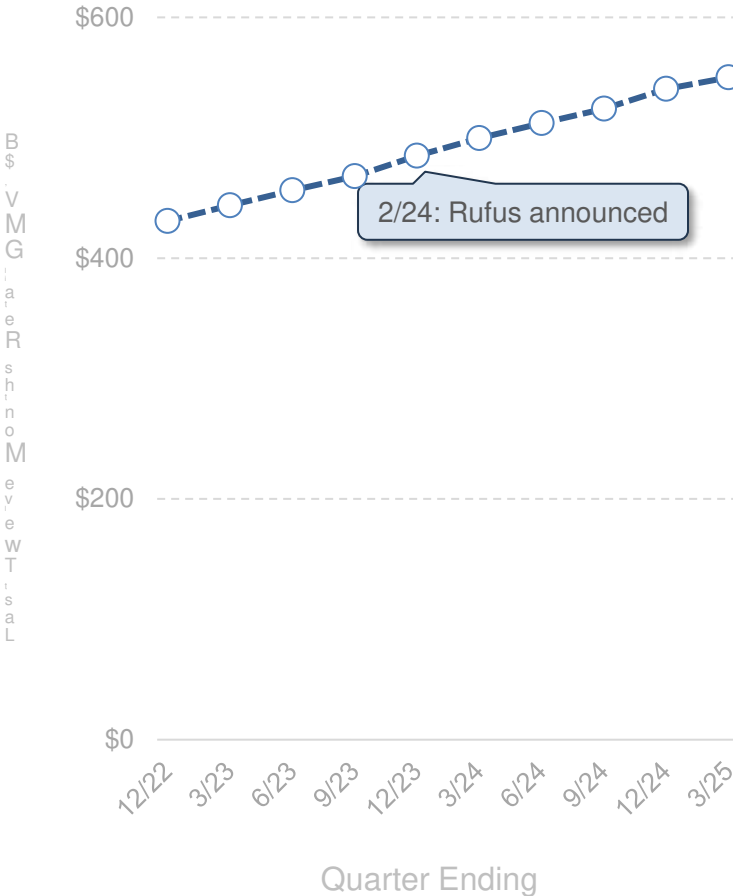


我们现在有很多客户使用 Rufus 来帮助他们快速找到有关产品的资料。他们还使用 Rufus 来总结客户评论，这样他们就不必阅读 100 条客户评论来了解人们对该产品的看法 …… 个性化程度越来越高 ……

…… 因此，我们预计在 2025 年全年，您不确定要购买什么，并且希望 Rufus 提供帮助的情况将会继续增加，并且对客户越来越有帮助。

- 亚马逊 CEO Andy Jassy, 2/25

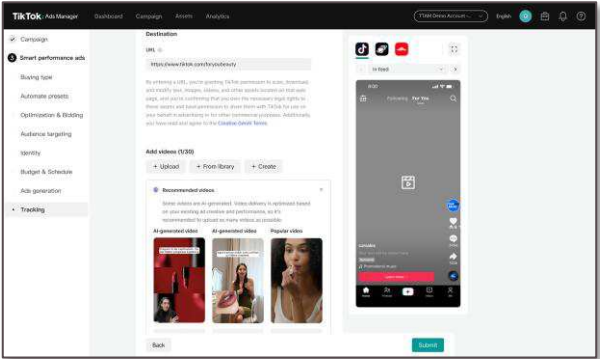
亚马逊北美零售预估商品总值（十亿美元），近 12 个月



来源：亚马逊；摩根士丹利估计

# Tech Incumbent AI Rollouts = TikTok – Symphony AI Assistant (6/24)

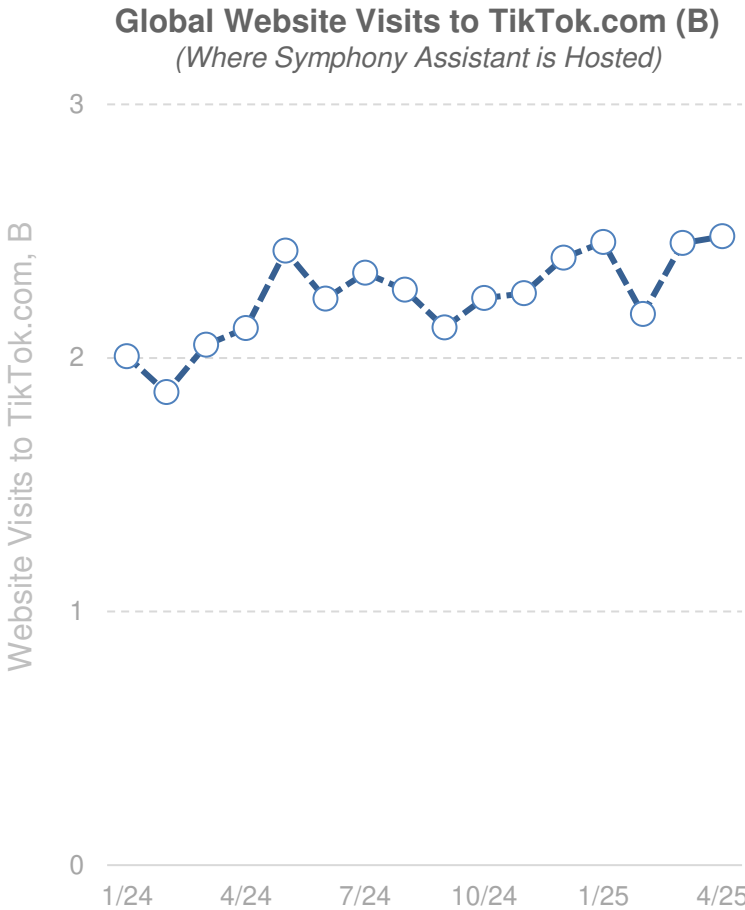
## TikTok: Symphony Assistant – 1/24-4/25, per TikTok & Similarweb



*Creativity thrives on TikTok. When brands truly lean into creative bravery and experimentation, they are able to speak directly to their community and invite them to join in the conversation. At TikTok World 2024 we launched Symphony, our suite of ad solutions powered by generative AI...*

*...With Symphony, businesses of all sizes, creators and agencies can blend human imagination with AI-powered efficiency to help scale content development, creativity, and productivity on TikTok. Research has proven that not only does creating TikTok-first ads boost purchase intent by +37% and brand favorability by +38%, but also 79% of TikTok users show a preference for brands that demonstrate a clear understanding of how to create content specifically for the platform.*

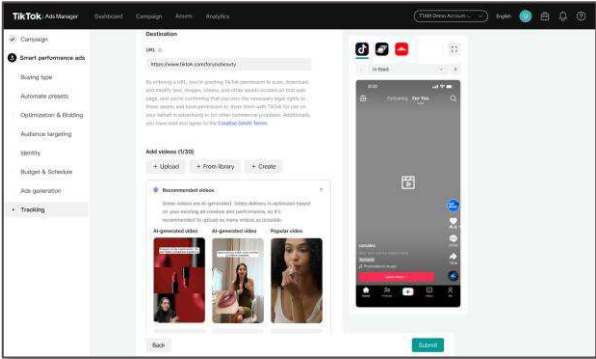
- TikTok Press Release, 6/24



Note: Includes both mobile & desktop website visits. China data may be subject to informational limitations due to government restrictions.  
Source: TikTok; Similarweb (5/25)

# 科技巨头 AI 推广 =TikTok – Symphony AI 助手 (6/24)

## TikTok: Symphony Assistant – 1/24-4/25, per TikTok & Similarweb



创意在 TikTok 上蓬勃发展。当品牌真正倾向于创造性的勇敢和实验时，他们可以直接与他们的社区对话，并邀请他们参与对话。在 TikTok World 2024 上，我们推出了 Symphony，我们由生成式 AI 驱动的广告解决方案套件 ...

... 借助 Symphony，各种规模的企业、创作者和代理商可以将人类的想象力与人工智能驱动的效率相结合，以帮助扩展 TikTok 上的内容开发、创造力和生产力。研究表明，创建 TikTok 优先广告不仅可以将购买意愿提高 +37%，品牌好感度提高 +38%，而且 79% 的 TikTok 用户更喜欢那些清楚地了解如何专门为该平台创建内容的品牌。

- TikTok 新闻稿，6/24

注意：包括移动和桌面网站访问。由于政府限制，中国数据可能受到信息限制。来源：TikTok; Similarweb (5/25)

Tech Incumbent AI Rollouts = Apple – Apple Intelligence (10/24)

Apple: Apple Intelligence – 9/23-3/25, per Apple & IDC Estimates

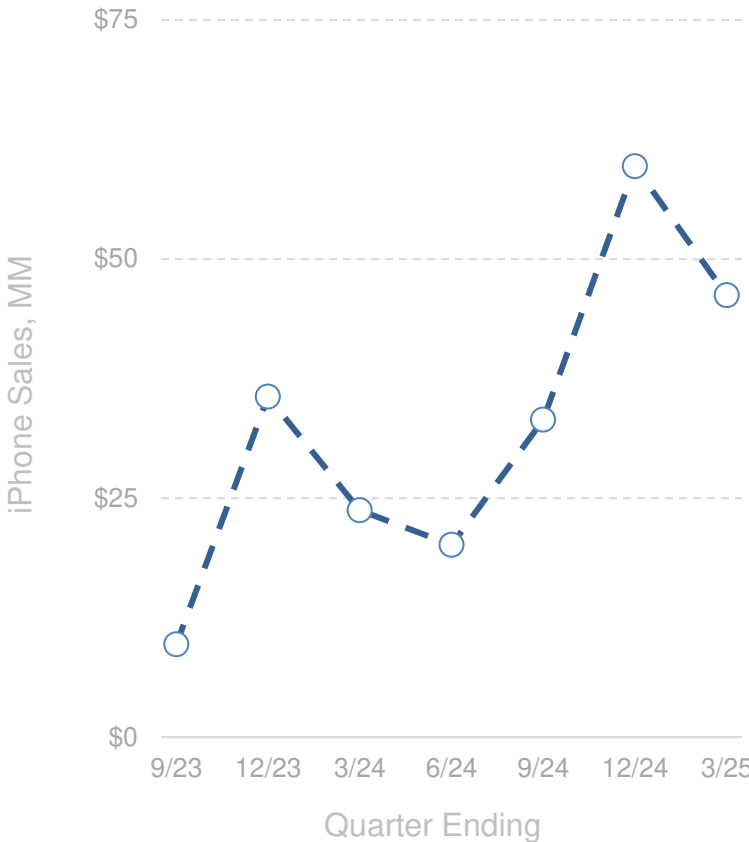


Apple Intelligence builds on years of innovations we’ve made across hardware and software to transform how users experience our products. Apple Intelligence also empowers users by delivering personal context that’s relevant to them. And importantly, Apple Intelligence is a breakthrough for privacy and AI with innovations like private cloud compute...

...[in] the markets where we had rolled out Apple Intelligence...year over year performance on the iPhone 16 family was stronger than those where Apple Intelligence was not available.

- Apple CEO Tim Cook, 1/25

Estimated Global Sales of iPhone 15 Pro / Pro Max & iPhone 16 (MM) – 9/23-3/25  
Apple Intelligence-Capable Devices



Note: Counts sales of iPhone 15 Pro, iPhone 15 Pro Max, & iPhone 16 devices. Figures are estimates.  
Source: Company announcements & investor filings; IDC via Morgan Stanley (4/25)

科技巨头人工智能发布 =Apple – Apple Intelligence (10/24)

Apple: Apple Intelligence – 9/23-3/25, 根据 Apple 和 IDC 估计

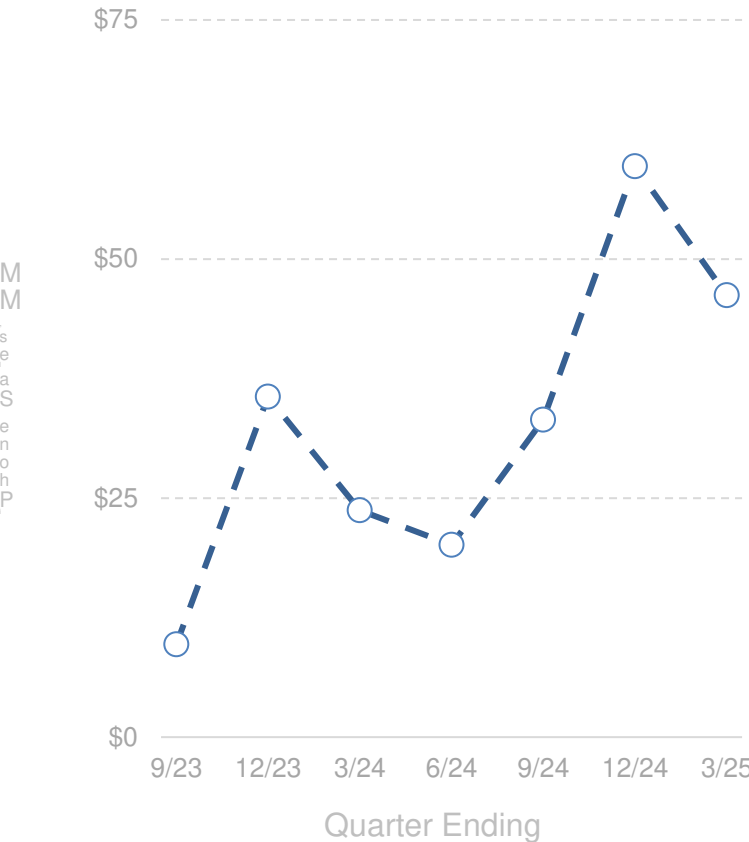


Apple Intelligence 基于我们多年来在硬件和软件方面的创新，旨在改变用户体验我们产品的方式。Apple Intelligence 还通过提供与用户相关的个人背景信息来增强用户能力。重要的是，Apple Intelligence 是隐私和人工智能方面的一项突破，它采用了诸如私有云计算之类的创新技术 ……

...[在] 我们推出 Apple Intelligence 的市场中 ……  
iPhone 16 的同比增长表现系列产品比未提供 Apple Intelligence 的产品更强。

- Apple CEO Tim Cook, 1/25

Estimated Global Sales of iPhone 15 Pro / Pro Max & iPhone 16 (MM) – 9/23-3/25  
Apple Intelligence-Capable Devices



注意：统计了 iPhone 15 Pro、iPhone 15 Pro Max 和 iPhone 16 设备的销量。数字为估计值。来源：公司公告和投资者备案；IDC（通过摩根士丹利）（4/25）

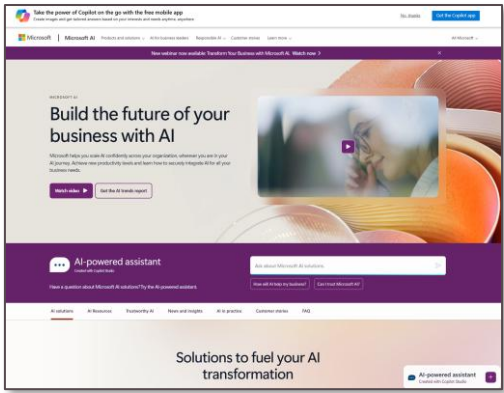
*AI – Tech Incumbents =*  
*Rapid Revenue + Customer Growth*

AI – 科技巨头 =  
收入快速增长 + 客户增长



# AI Monetization – ‘AI Product’ = Microsoft AI Revenue +175% to \$13B Y/Y

## Microsoft AI Product Revenue – 2023-2024, per Microsoft

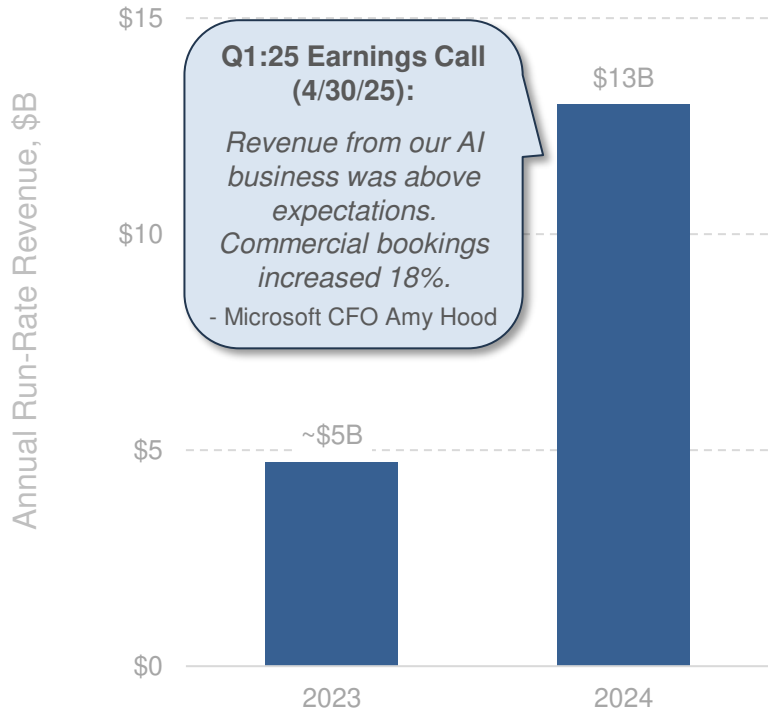


We are innovating across our tech stack and helping customers unlock the full ROI of AI to capture the massive opportunity ahead...

...Already, our AI business has surpassed an annual revenue run rate of \$13 billion, up 175% year-over-year.

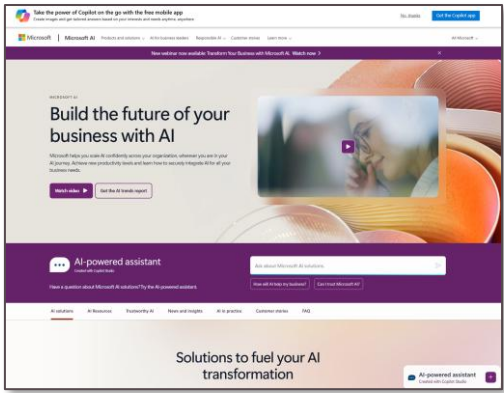
- Microsoft CEO Satya Nadella, 1/25

### Estimated Microsoft AI Product Annual Run-Rate Revenue (\$B)



# AI Monetization – ‘AI Product’ = Microsoft AI Revenue +175% to \$13B Y/Y

## Microsoft AI 产品收入 – 2023-2024, 数据来源：微软

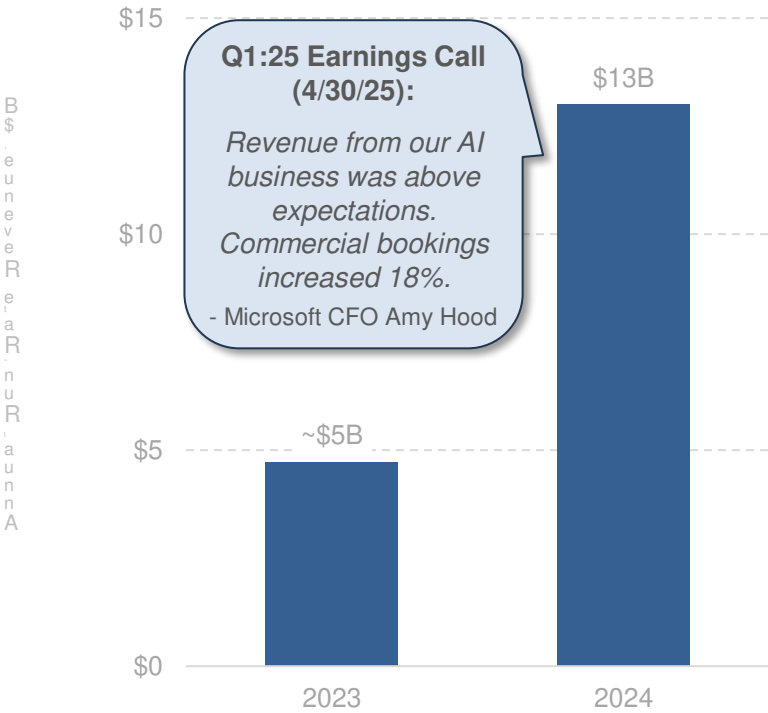


我们正在整个技术堆栈中进行创新，并帮助客户充分释放 AI 的投资回报率，以抓住未来的巨大机遇

…… 目前，我们的 AI 业务年收入运行率已超过 130 亿美元，同比增长 175%。

- 微软 CEO Satya Nadella, 1/25

### 微软 AI 产品年度营收预估（10 亿美元）

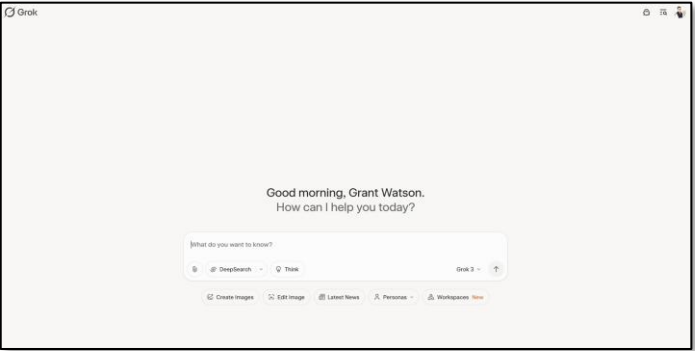


Note: Microsoft AI revenue likely includes Azure AI services, Microsoft 365 Copilot, GitHub Copilot, Dynamics 365 Copilot, Azure OpenAI Services, and others. Detailed breakdowns not provided on earnings calls. Source: Microsoft Press Release, 'Microsoft Cloud and AI strength drives second quarter results' (1/25); & other Microsoft announcements

注意：微软 AI 收入可能包括 Azure AI 服务、Microsoft 365 Copilot、GitHub Copilot、Dynamics 365 Copilot、Azure OpenAI Services 等。财报电话会议未提供详细分类。来源：微软新闻稿“Microsoft Cloud and AI strength drives second quarter results”（1/25）；& 其他微软公告

# AI Monetization – Generative Search = xAI Annualized Revenue Up Materially in 2025

## xAI: Generative Search, per xAI & The Wall Street Journal

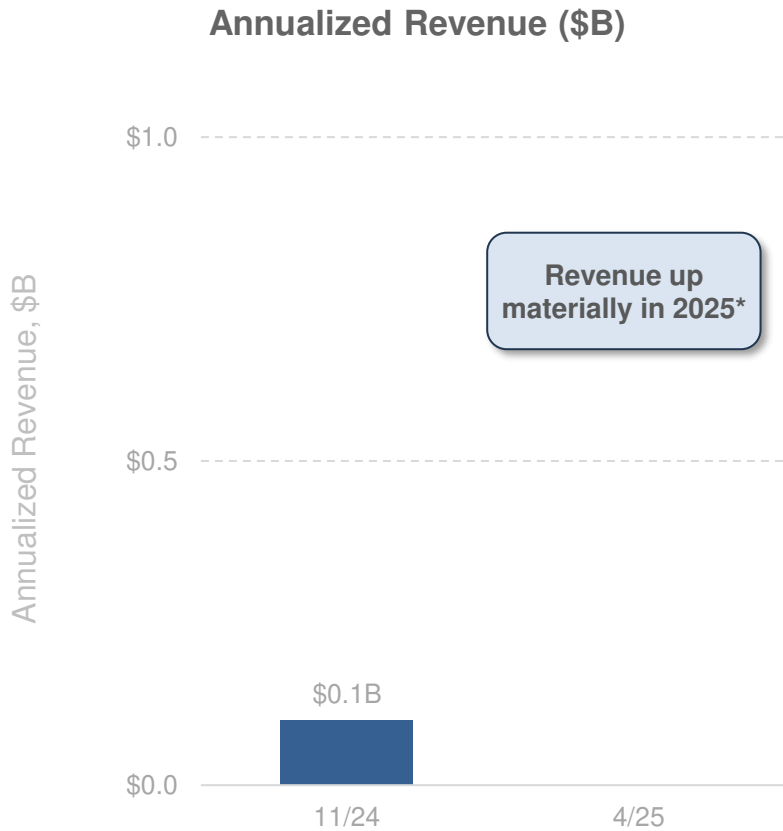


We are pleased to introduce Grok 3, our most advanced model yet: blending strong reasoning with extensive pretraining knowledge. Trained on our Colossus supercluster with 10x the compute of previous state-of-the-art models, Grok 3 displays significant improvements in reasoning, mathematics, coding, world knowledge, and instruction-following tasks.

- xAI Grok 3 Press Release, 2/25

[Grok is a] maximally truth-seeking AI, even if that truth is sometimes at odds with what is politically correct.

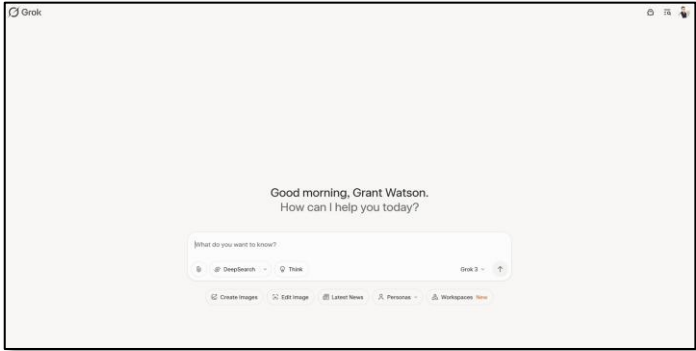
- xAI Founder & CEO Elon Musk, 2/25



\*Select media reports have xAI revenue being as high as \$1B as of 4/25. Source: xAI (2/25); The Wall Street Journal, 'Elon Musk's xAI Startup Is Valued at \$50 Billion in New Funding Round' (11/24) ([link](#)); CNBC, 'Musk says he's looking to put 'proper value' on xAI during investor call, sources say' (4/25) ([link](#))

# AI Monetization – Generative Search = xAI Annualized Revenue Up Materially in 2025

## xAI：生成式搜索，根据 xAI 和《华尔街日报》报道

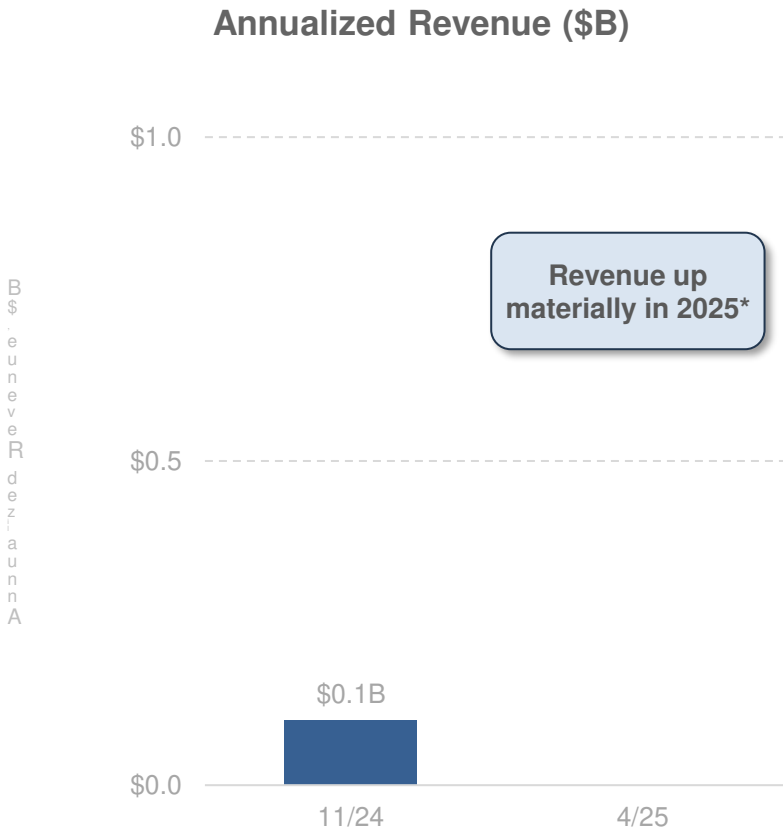


We are pleased to introduce Grok 3, our most advanced model yet: blending strong reasoning with extensive pretraining knowledge. Trained on our Colossus supercluster with 10x the compute of previous state-of-the-art models, Grok 3 displays significant improvements in reasoning, mathematics, coding, world knowledge, and instruction-following tasks.

- xAI Grok 3 Press Release, 2/25

[Grok is a] maximally truth-seeking AI, even if that truth is sometimes at odds with what is politically correct.

- xAI Founder & CEO Elon Musk, 2/25



\*Select media reports have xAI revenue being as high as \$1B as of 4/25. Source: xAI (2/25); The Wall Street Journal, 'Elon Musk's xAI Startup Is Valued at \$50 Billion in New Funding Round' (11/24) ([link](#)); CNBC, 'Musk says he's looking to put 'proper value' on xAI during investor call, sources say' (4/25) ([link](#))



# AI Monetization – AI Services = Palantir USA Commercial Customers +65% to 432 Y/Y

Palantir USA Commercial Customers – Q1:23-Q1:25, per Palantir



We achieved a \$1 billion annual run rate in our US commercial business for the first time as AIP [Artificial Intelligence Platform] continues to drive both new customer conversions and existing customer expansions in the US..

- Palantir CFO David Glazer , 5/25

As AI models progress and improve, we continue enabling our customers to maximally leverage these models in production, capitalizing upon the rich context within the enterprise through the Ontology. We remain differentiated in our elite execution to deliver quantified exceptionalism for our customers, ever widening their advantage over the AI have-nots.

- Palantir CRO & Chief Legal Officer Ryan Taylor, 5/25

Source: Palantir

# AI 货币化 – AI 服务 =Palantir 美国商业客户 +65% 至 432 Y/Y

Palantir 美国商业客户 – Q1:23-Q1:25, 数据来源：Palantir



由于 AIP [人工智能平台] 不断推动美国新客户转化和现有客户扩张，我们的美国商业业务首次实现了 10 亿美元的年度运行率。

- Palantir 首席财务官 David Glazer, 5/25

随着 AI 模型的进步和改进，我们将继续使我们的客户能够在生产中最大限度地利用这些模型，通过本体利用企业内部丰富的上下文。我们仍然以卓越的执行力脱颖而出，为我们的客户提供可量化的卓越性，不断扩大他们在人工智能领域的优势。

- Palantir CRO & Chief Legal Officer Ryan Taylor, 5/25

来源：Palantir

AI Monetization Possibilities – Enterprise =

Horizontal Platform  
& / Or  
Specialized Software?

AI 货币化可能性 – 企业 =

横向平台和 / 或专  
业软件?

To understand where enterprise AI monetization is headed, it helps to ask where software itself is consolidating.

For decades, business software followed a familiar pattern: build a specialized tool, sell it to a narrow user base, and scale up within a vertical. This was the age of vertical SaaS – Toast for restaurants, Guidewire for insurance, Veeva for life sciences...Each tool solved a deep, narrow problem. But with the rise of foundation models and generative AI, others are gunning for these prizes.

Enter the horizontal enterprise platforms – horizontal layers that combines AI-native productivity, search, communication, and knowledge management into one unified interface. Think of it as Slack meets Notion meets ChatGPT, all in one platform.

Horizontal enterprise platforms could usher in a new form of monetization: not by selling siloed software licenses, but by charging for intelligence, embedded throughout the stack. The value shifts from tools to outcomes – from CRMs to automated deal summaries, from service desks to AI-powered resolution flows.

These horizontal capabilities are still early, but they're already being harnessed by incumbents and upstarts alike. Microsoft is integrating Copilot across the stack. Zoom and Canva are layering GenAI into user-facing workflows, while Databricks is infusing GenAI into its data and developer stack. Meanwhile, startups like Glean are betting on AI-first workflows to challenge the suite model...

要了解企业人工智能货币化的发展方向，首先需要了解软件本身的整合方向。

几十年来，商业软件遵循着一种熟悉的模式：构建专门的工具，将其出售给狭窄的用户群，并在垂直领域内进行扩展。这是垂直 SaaS 的时代 – 餐厅的 Toast、保险的 Guidewire、生命科学的 Veeva…… 每种工具都解决了深刻而狭窄的问题。但随着基础模型和生成式 AI 的兴起，其他人也开始争夺这些奖励。

进入横向企业平台 – 将 AI 原生生产力、搜索、通信和知识管理结合到一个统一界面中的横向层。可以将其视为 Slack、Notion 和 ChatGPT 的结合体，全部集成在一个平台中。

横向企业平台可能会迎来一种新的货币化形式：不是通过销售孤立的软件许可证，而是通过对嵌入到整个堆栈中的智能收费。价值从工具转移到结果 – 从 CRM 到自动交易摘要，从服务台到 AI 驱动的解决方案流程。

这些横向功能仍处于早期阶段，但它们已经被现有企业和新兴企业所利用。微软正在将 Copilot 集成到整个堆栈中。Zoom 和 Canva 正在将 GenAI 分层到面向用户的工作流程中，而 Databricks 正在将 GenAI 注入到其数据和开发者堆栈中。与此同时，Glean 等初创公司正在押注 AI-first workflows to challenge the suite model...



...But specialist vendors aren't standing still. If anything, they're absorbing AI faster – embedding copilots, automating workflows, and fine-tuning models on proprietary industry data. These platforms already have the workflows, the trust, and the structured data that AI thrives on. That gives them a head start in deploying domain-specific intelligence – AI that doesn't just summarize a meeting, but flags regulatory risks, optimizes pricing in real time, or drafts FDA-compliant documentation. In many cases, their incumbency becomes their advantage: they can roll out AI as a feature, not a product, and monetize it without changing the buying motion.

The next chapter of AI monetization may not be a winner-take-all battle, but a convergence. Horizontal platforms will push breadth, stitching together knowledge across functions; specialists will push depth, delivering AI that speaks the language of compliance, contracts, and customer intent.

The question isn't whether platforms or specialists win – it's who can abstract the right layer, own the interface, and capture the logic of work itself. In the AI era, monetization won't just follow usage – it will follow attention, context, and control.

... 但是，专业供应商并没有停滞不前。如果有什么不同的话，那就是他们吸收人工智能的速度更快 —— 嵌入副驾驶，自动化工作流程，并根据专有的行业数据微调模型。这些平台已经拥有人工智能赖以生存的工作流程、信任和结构化数据。这使他们在部署特定领域的智能方面具有领先优势 —— 人工智能不仅可以总结会议，还可以标记监管风险，实时优化定价，或起草符合 FDA 标准的文档。在许多情况下，他们现有的地位变成了他们的优势：他们可以将人工智能作为一项功能而非产品推出，并在不改变购买方式的情况下将其货币化。

人工智能货币化的下一篇章可能不是一场赢者通吃的战斗，而是一场融合。横向平台将推动广度，将跨职能的知识缝合在一起；专业平台将推动深度，提供能够理解合规性、合同和客户意图的人工智能。

问题不在于平台或专家是否获胜 —— 跟踪正确的层，拥有 it's who can abstract the right layer, own the interface, and capture the logic of work itself. In the AI era, monetization won't just follow usage – it will follow attention, context, and control.

AI Monetization Possibilities – Enterprise =

**Horizontal Platform**  
& / Or  
Specialized Software?

AI 货币化可能性 – 企业 =

横向平台  
& / 或  
Specialized Software?

*Horizontal Enterprise Platform =*

***SaaS Incumbents***

*Or*

*Large Language Model Challengers?*

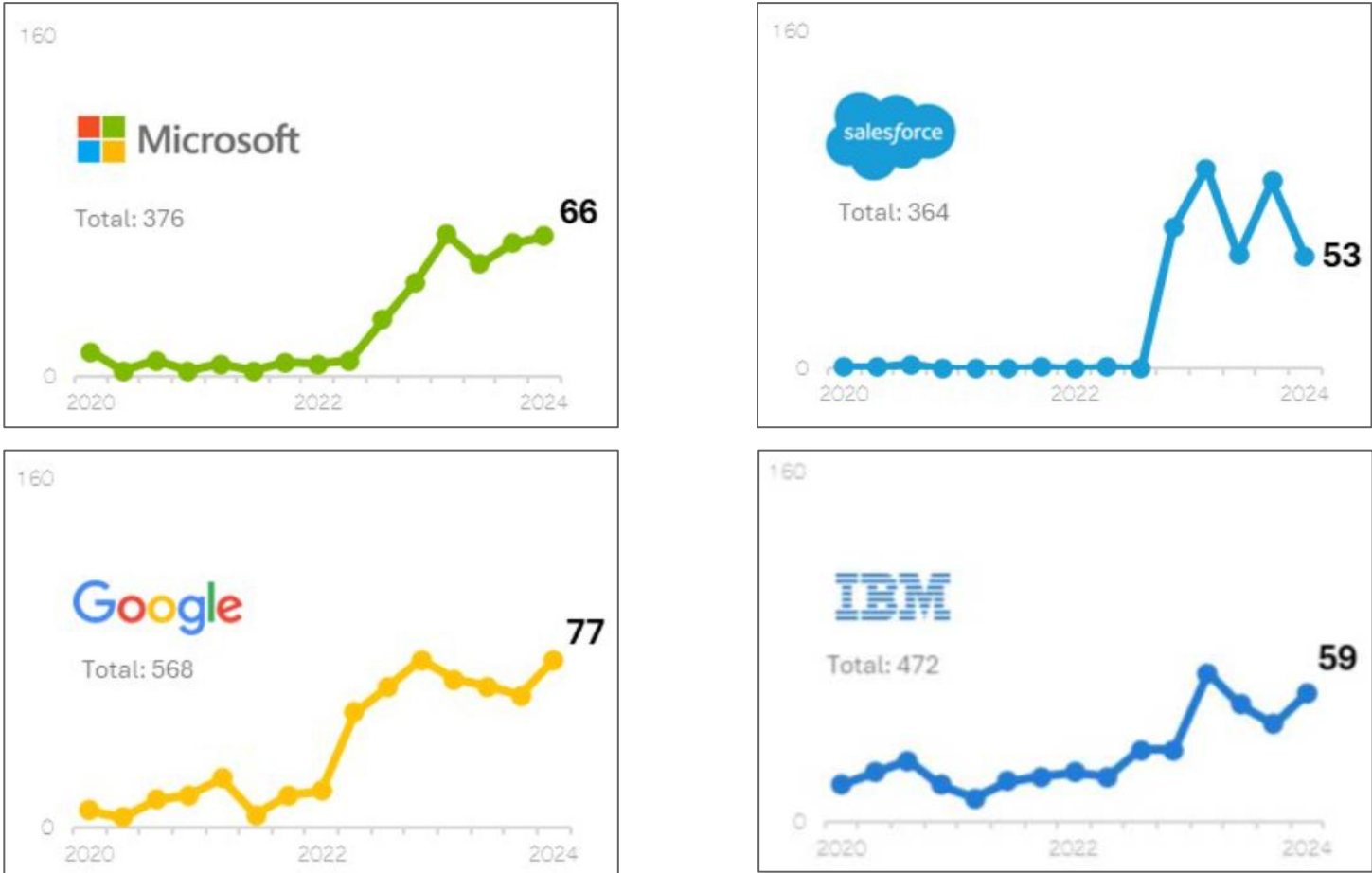
水平企业平台 =

SaaS 现有企业或

大型语言模型挑战者？

# Enterprise SaaS Incumbent AI Rollouts = Broad & Steady Cadence

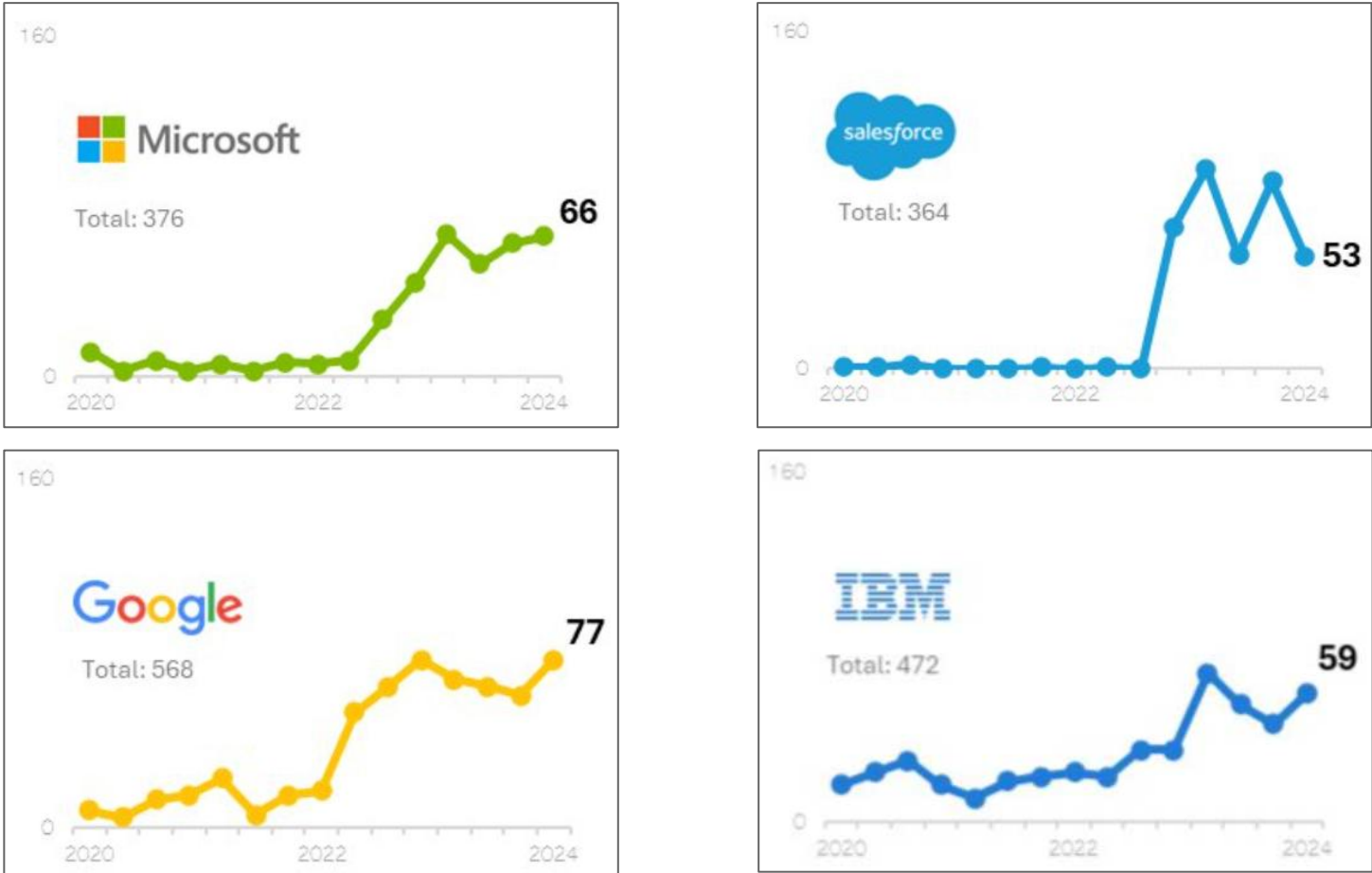
Number of Mentions of ‘AI’ on Corporate Earnings Calls – Q1:20-Q1:24, per Uptrends.ai



Source: Uptrends.ai (6/24), company announcements & investor filings

# Enterprise SaaS Incumbent AI Rollouts = Broad & Steady Cadence

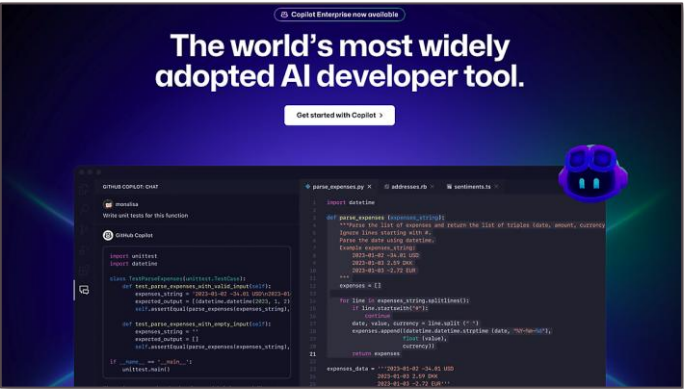
公司财报电话会议中 “AI” 的提及次数 – Q1:20-Q1:24, per Uptrends.ai



来源: Uptrends.ai (6/24), 公司公告和投资者备案

Enterprise SaaS Incumbent AI Rollouts = 企业 SaaS 现有 AI 推广 =Microsoft GitHub Copilot – 6/22

Microsoft GitHub Copilot – 6/17-6/24, per GitHub, Microsoft & Wells Fargo

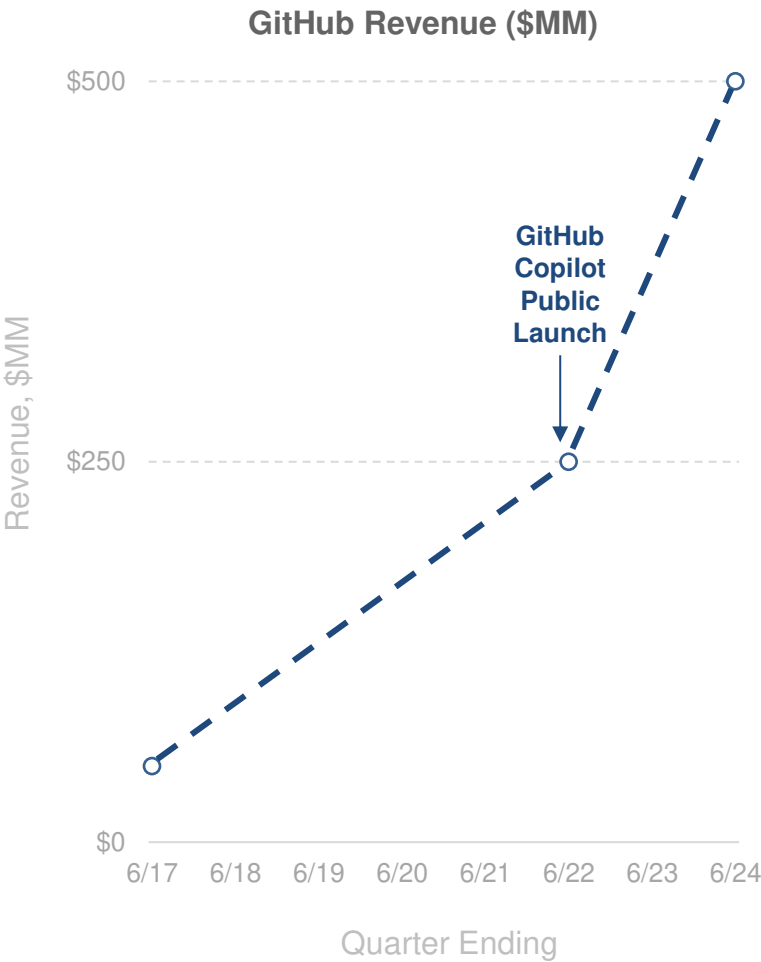


GitHub Copilot is by far the most widely adopted AI-powered developer tool. Just over two years since its general availability, more than 77,000 organizations – from BBVA, FedEx, and H&M, to Infosys and Paytm – have adopted Copilot, up 180% year-over-year.

- Microsoft CEO Satya Nadella, 7/24

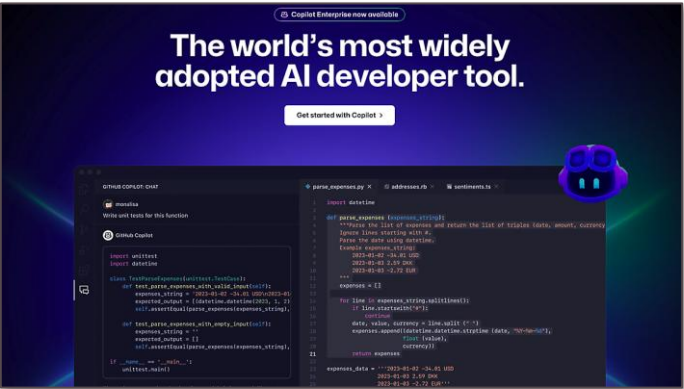
We have been delighted by the early response to GitHub Copilot and vs. Code with more than 1 million sign-ups in just the first week post launch. All up, GitHub now is home to 150 million developers, up 50% over the past two years.

- Microsoft CEO Satya Nadella, 1/25



Note: GitHub revenue is disclosed irregularly; 3 datapoints are from company leadership's disclosures. Public developer launch date shown. GitHub reports annualized revenue; here, we translate this to quarterly revenue. Source: Company announcements & investor filings

Microsoft GitHub Copilot – 6/17-6/24, 根据 GitHub、Microsoft 和 Wells Fargo

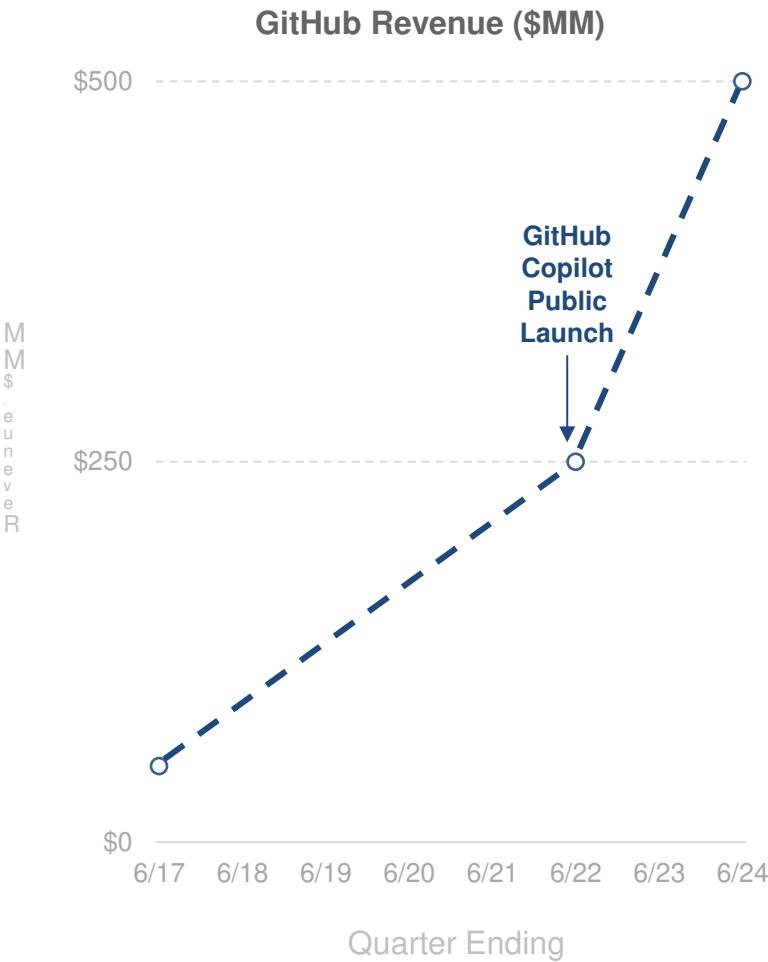


GitHub Copilot 是迄今为止应用最广泛的 AI 驱动的开发工具。自其普遍可用性以来仅仅两年多，超过 77,000 家组织 – 从 BBVA、FedEx 和 H&M 到 Infosys 和 Paytm – 采用了 Copilot，同比增长 180%。

- Microsoft CEO Satya Nadella, 7/24

GitHub Copilot 和 vs. Code 在发布后的第一个星期内就获得了超过 100 万的注册用户，我们对此感到非常高兴。目前，GitHub 拥有 1.5 亿开发者，在过去两年中增长了 50%。

- Microsoft CEO Satya Nadella, 1/25



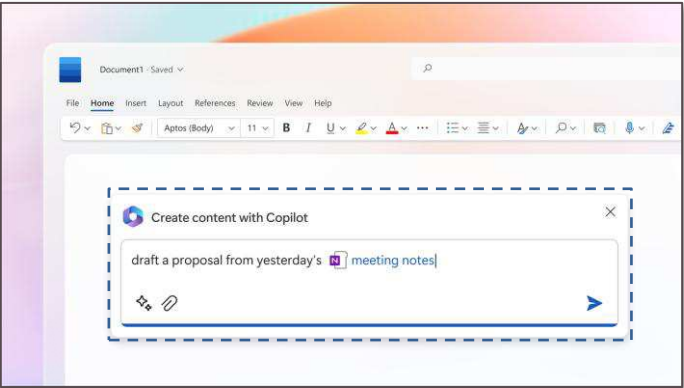
注意：GitHub 的收入披露不规律；3 个数据点来自公司领导层的披露。显示了公共开发者 launch 日期。GitHub 报告年度收入；这里，我们将其转换为季度收入。来源：公司公告和投资者备案



Enterprise SaaS Incumbent AI Rollouts = Microsoft 365 Copilot – 3/23

Enterprise SaaS Incumbent AI Rollouts = Microsoft 365 Copilot – 3/23

Microsoft 365 Copilot – Q2:23-Q4:24, per Microsoft & Morgan Stanley



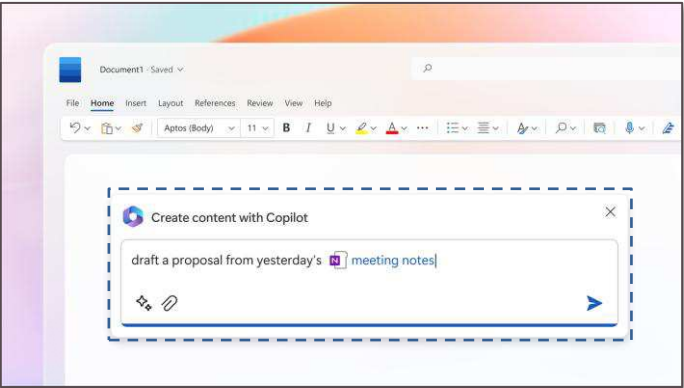
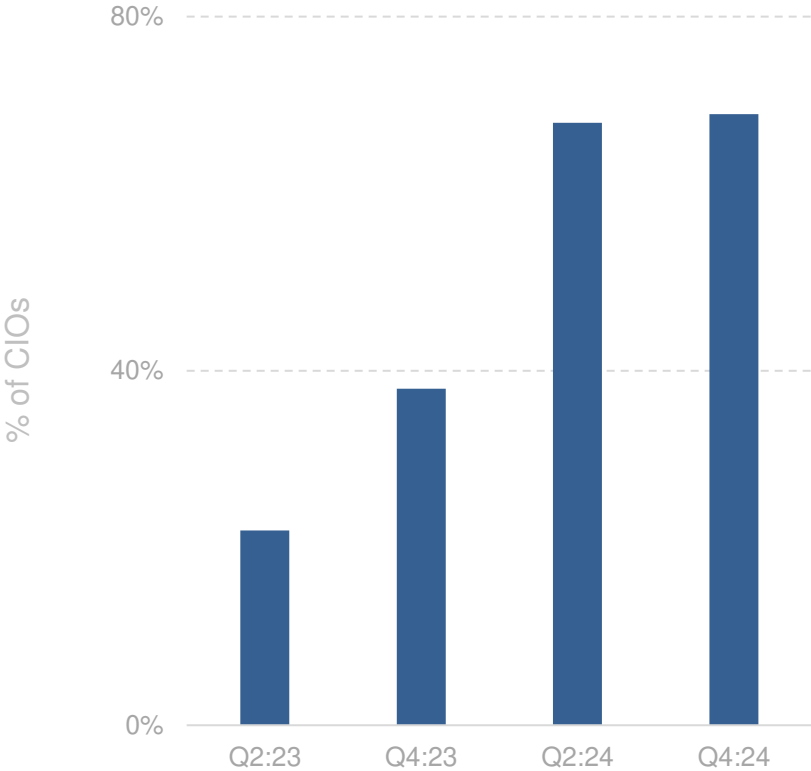
We are seeing accelerated customer adoption across all deal sizes as we win new Microsoft 365 Copilot customers and see the majority of existing enterprise customers come back to purchase more seats. When you look at customers who purchased Copilot during the first quarter of availability, they have expanded their seat collectively by more than 10x over the past 18 months. And overall, the number of people who use Copilot daily, again, more than doubled quarter over quarter.

Employees are also engaging with Copilot more than ever. Usage intensity increased more than 60% quarter over quarter, and we are expanding our TAM with Copilot Chat, which was announced earlier this month.

- Microsoft CEO Satya Nadella, 1/25

Note: N=61 CIOs in the USA & EU. Microsoft 365 Copilot was announced in 3/23 but was not made generally available for enterprise customers until 11/23. Source: Company announcements & investor filings, Morgan Stanley, '4Q24 Preview – Can Microsoft Add Clarity to the AI Monetization Question?' (7/24)

% of CIOs Expecting to Use Microsoft 365 Copilot over Next 12 Months, per Morgan Stanley Survey



我们看到所有规模的交易中客户的采用都在加速，因为我们赢得了新的 Microsoft 365 Copilot 客户，并且看到大多数现有企业客户回来购买更多席位。当您查看在首次可用季度购买 Copilot 的客户时，他们在过去 18 个月中将其席位总共扩大了 10 倍以上。总的来说，每天使用 Copilot 的人数，再次，环比增长了一倍以上。

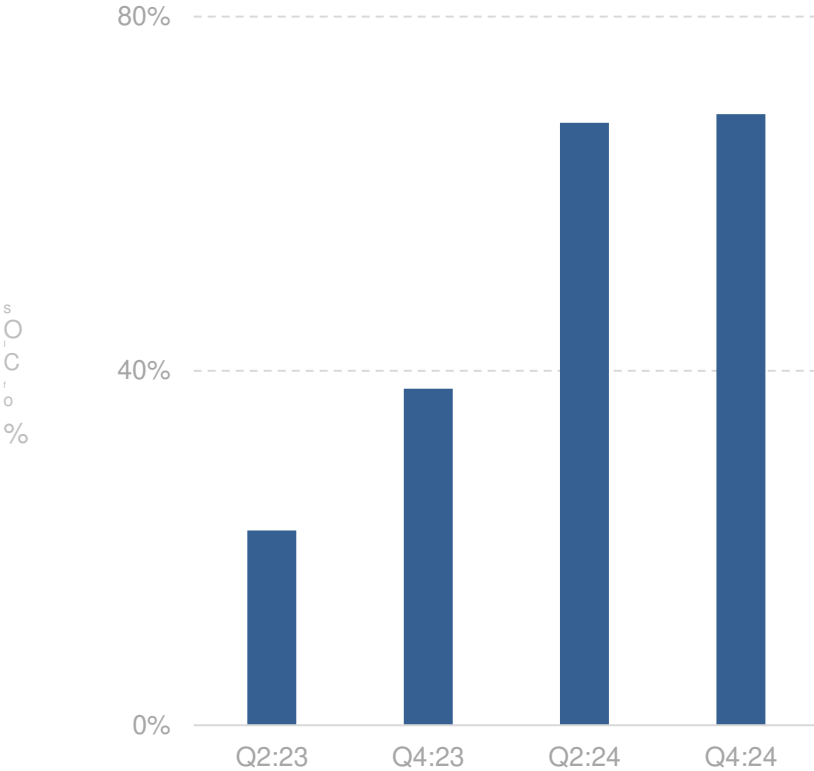
员工对 Copilot 的参与度也比以往任何时候都高。使用强度环比增长超过 60%，我们正在通过本月早些时候宣布的 Copilot Chat 扩大我们的 TAM。

- Microsoft CEO Satya Nadella, 1/25

注意：美国和欧盟的 N=61 CIO。Microsoft 365 Copilot 于 3/23 发布，但直到 11/23 才向企业客户普遍提供。来源：公司公告和投资者备案，摩根士丹利，"4Q24 预览 – 微软能否澄清 AI 货币化问题？"(7/24)

Microsoft 365 Copilot – Q2:23-Q4:24, 根据微软和摩根士丹利的数据

根据摩根士丹利调查，未来 12 个月内预计使用 Microsoft 365 Copilot 的首席信息官百分比



Adobe Firefly – 5/23-4/25, per Adobe

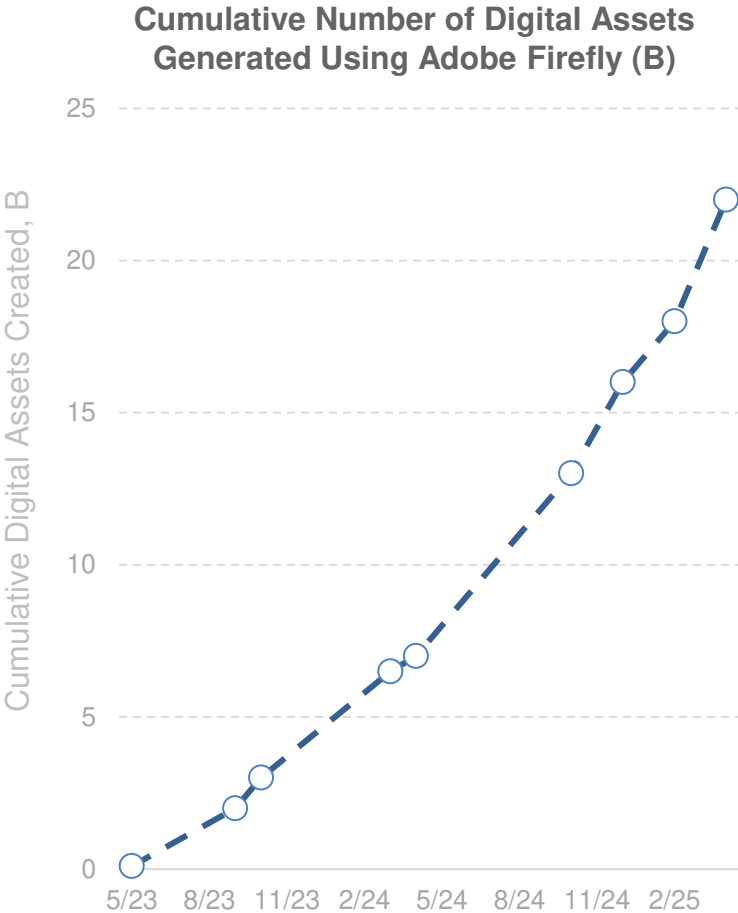


The release of the Adobe FireFly video model in February, a commercially safe generative AI video model, has been very positively received by brands and creative professionals...

...User engagement has been strong with over 90% of paid users generating videos...

...We're delighted with the early interest in these new offerings. Other creative professional and creator highlights include, continued strong adoption of GenAI in our products with Photoshop GenAI monthly active users at approximately 35% and Lightroom GenAI monthly active users at 30%. Users have generated over 20 billion assets with Firefly.

- Adobe President of Digital Media David Wadhwani, 3/25



Note: We assume zero users in the launch month. Adobe Firefly was released as a public beta in March 2023.  
Source: Adobe announcements (9/23, 10/23, 3/24, 4/24, 10/24, 12/24, 2/25)

Adobe Firefly – 5/23-4/25, per Adobe



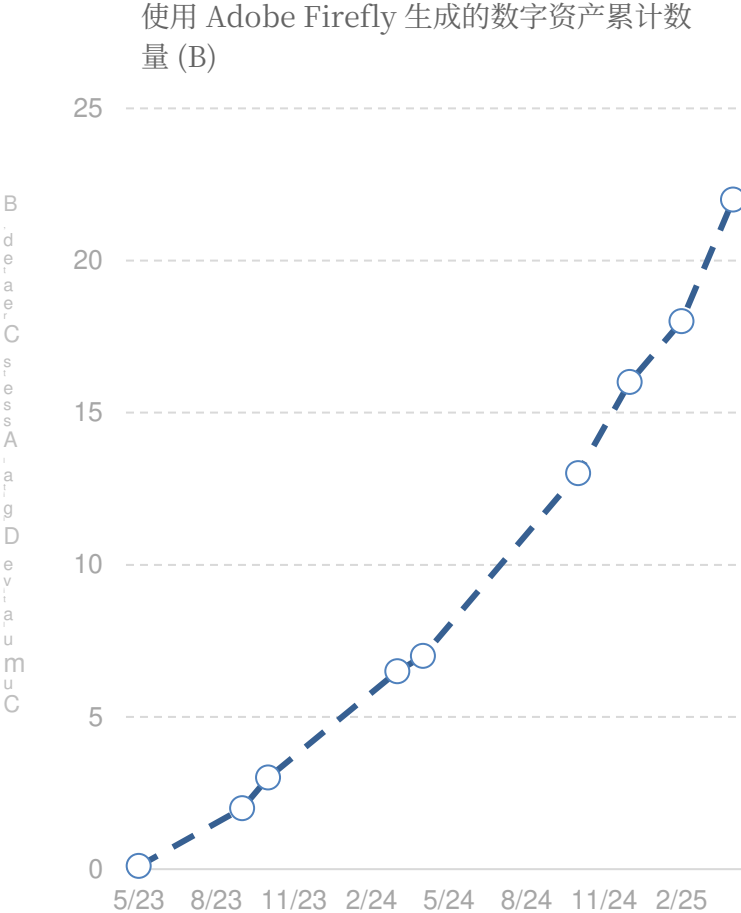
2 月份发布的 Adobe FireFly 视频模型是一种商业上安全的生成式 AI 视频模型，受到了品牌和创意专业人士的非常积极的评价 ...

... 用户参与度一直很高，超过 90% 的付费用户都在生成视频 ...

... 我们对这些新产品的早期兴趣感到高兴。其他创意专业人士和创作者的亮点包括，我们的产品继续大力采用 GenAI， Photoshop GenAI 的月活跃用户约为 35%， Lightroom GenAI 的月活跃用户为 30%。用户已经使用 Firefly 生成了超过 200 亿个资产。

- Adobe 数字媒体总裁 David Wadhwani, 3/25

注意：我们假设发布月份的用户数为零。Adobe Firefly 于 2023 年 3 月发布为公开测试版。来源：Adobe 公告 (9/23, 10/23, 3/24, 4/24, 10/24, 12/24, 2/25)



Enterprise SaaS Incumbent AI Rollouts =  
Atlassian Intelligence – 4/23

Enterprise SaaS Incumbent AI Rollouts =  
Atlassian Intelligence – 4/23

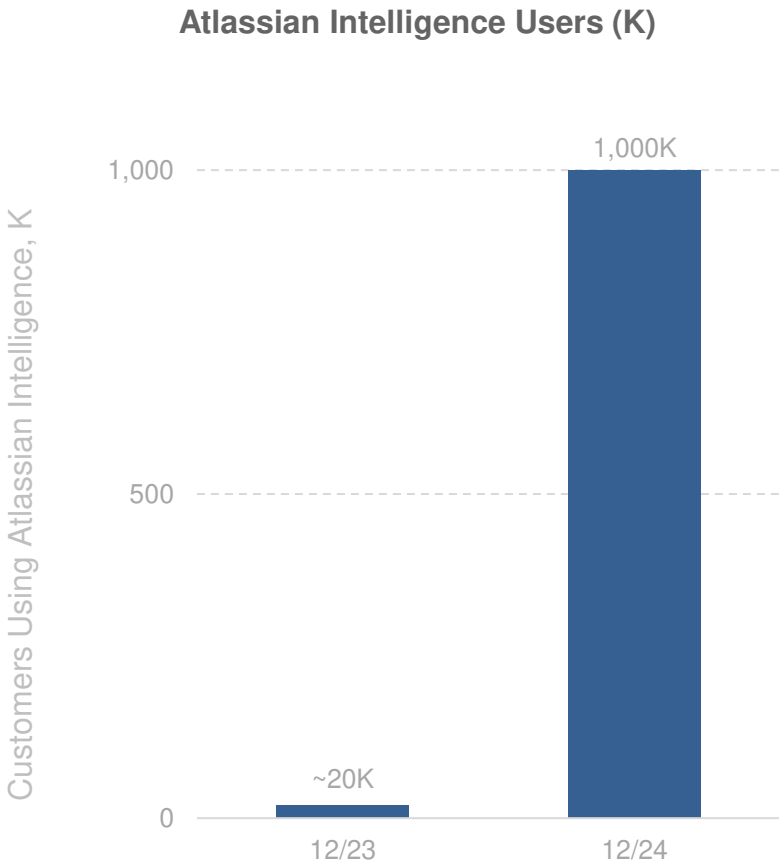
Atlassian Intelligence – 12/23-12/24, per Atlassian



Today, more than 1 million monthly active users are utilizing our Atlassian intelligence features to unlock enterprise knowledge, supercharge workflows, and accelerate their team collaboration. These features are clearly delivering value as we've seen a number of AI interactions increase more than 25x year over year...

...Atlassian Intelligence [saw a] 25x improvement in the number of features used over the last year.

- Atlassian Co-Founder & Co-CEO Michael Cannon, 2/25



Note: 12/23 users includes beta users. We assume 20,000 users based on Atlassian's disclosure that 'Nearly 10% of Atlassian's 265,000+ customers have already leveraged Atlassian Intelligence through our beta program.' Source: Atlassian announcements (4/23, 12/23, 12/24)

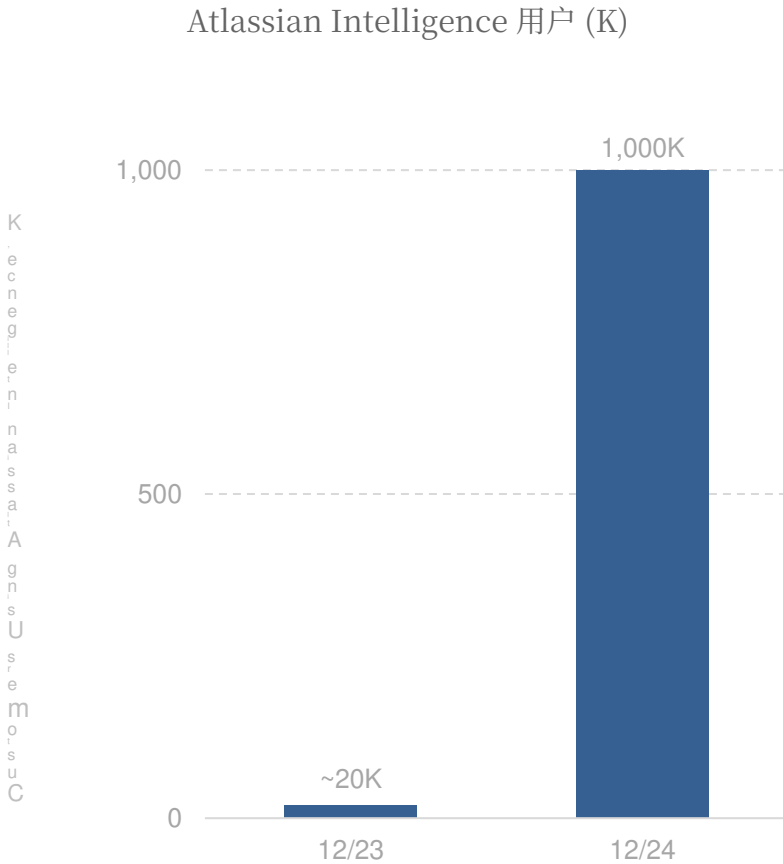
Atlassian Intelligence – 12/23-12/24, 根据 Atlassian。



如今，超过 100 万月活跃用户正在利用我们的 Atlassian intelligence 功能来解锁企业知识、增强工作流程并加速团队协作。这些功能显然在传递价值，因为我们已经看到许多 AI 互动同比增长超过 25 倍 ……

……Atlassian Intelligence [在过去一年中，使用的功能数量增加了] 25 倍。

- Atlassian 联合创始人兼联席 CEO Michael Cannon, 2/25



注意：12/23 的用户包括 beta 用户。我们假设有 20,000 名用户，这是基于 Atlassian 披露的“近 10% 的 Atlassian 的 265,000+ 客户已经通过我们的 beta 计划利用了 Atlassian Intelligence。”来源：Atlassian 公告（4/23、12/23、12/24）

Enterprise SaaS Incumbent AI Rollouts = Zoom AI Companion – 9/23

Enterprise SaaS Incumbent AI Rollouts = Zoom AI Companion – 9/23

Zoom AI Companion – 9/23-12/24, per Zoom

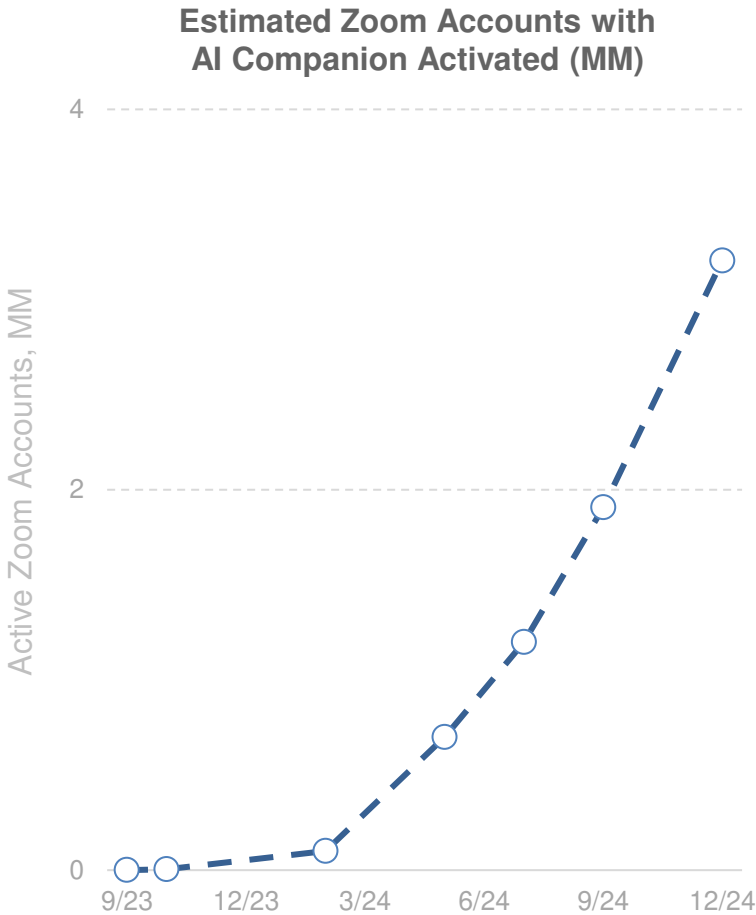
Zoom AI Companion – 9/23-12/24, per Zoom



Growth in monthly active users of Zoom AI Companion accelerated to 68% quarter over quarter, demonstrating the real value AI is providing customers. Zoom AI Companion has emerged as a driving force behind our transformation into an AI-first company...

...As part of AI Companion 2.0, we added advanced agentic capabilities, including memory, reasoning, orchestration, and seamless integration with Microsoft and Google services.

- Zoom Founder & CEO Eric Yuan, 2/25



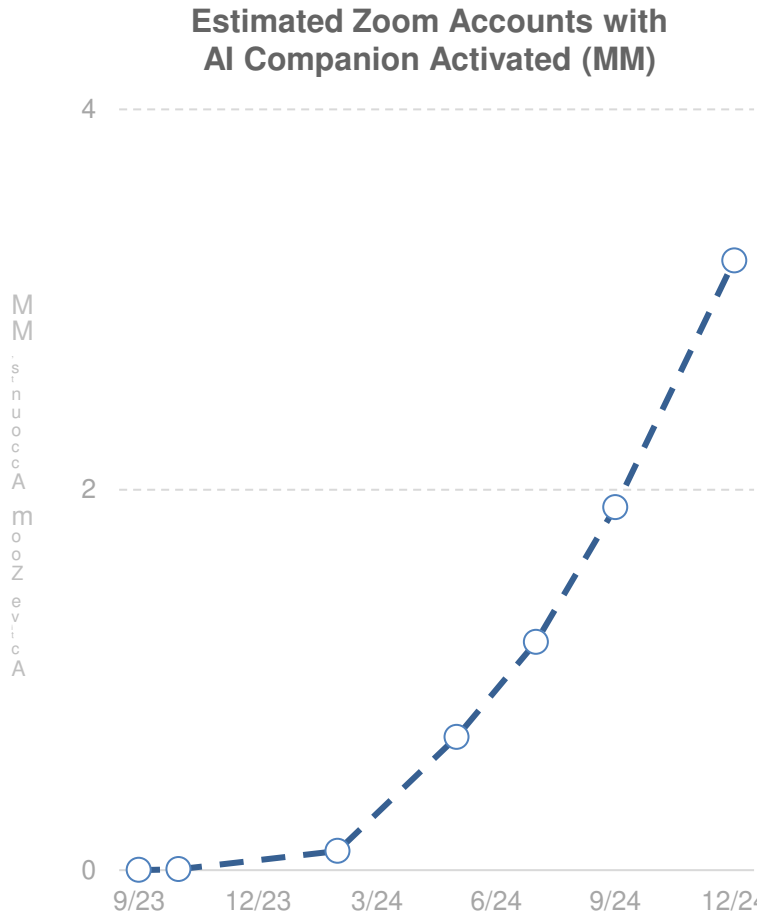
Note: AI Companion MAUs are estimates based on company disclosures. As of 7/30/24, Zoom disclosed they had 1.2MM accounts with AI Companion activated. In Q3 2024, they disclosed 59% Q/Q growth in active accounts; in Q4 2024, they disclosed further 68% Q/Q growth. We assume zero users in the launch month. Source: Zoom announcements (9/23, 10/23, 2/24, 5/24, 7/24, 9/24, 12/24)



Zoom AI Companion 的月活跃用户增长加速至季度环比增长 68%，表明 AI 为客户提供的真正价值。Zoom AI Companion 已成为我们转型为 AI-*first company*... 的驱动力。

...As part of AI Companion 2.0, we added advanced agentic 功能, 包括记忆、推理、协调以及与 Microsoft 和 Google 服务的无缝集成。

- Zoom 创始人兼 CEO Eric Yuan, 2/25



注意：AI Companion MAU 是基于公司披露的估计值。截至 2024 年 7 月 30 日，Zoom 披露他们有 120 万个已激活 AI Companion 的帐户。在 2024 年第三季度，他们披露活跃帐户环比增长 59%；在 2024 年第四季度，他们披露进一步环比增长 68%。我们假设发布月份的用户数为零。来源：Zoom 公告（9/23, 10/23, 2/24, 5/24, 7/24, 9/24, 12/24）

# Enterprise SaaS Incumbent AI Rollouts = Canva Magic Studio – 10/23

## Canva Magic Studio – 10/23-5/25, per Canva

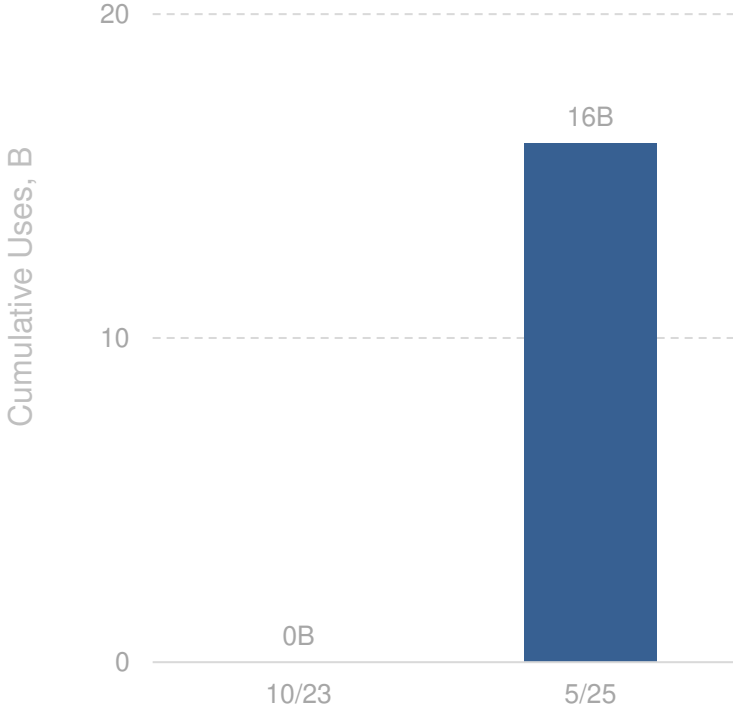


*With Magic Studio there's no need to toggle between multiple AI tools or learn lots of different software – all the best of AI is at your fingertips. Created for the 99% of the world without complex design skills, Magic Studio is jam-packed with easy-to-use AI-powered features across every part of Canva to help you work smarter.*

**- Canva Press Release, 10/23**

*Magic Studio is designed to supercharge creativity across our entire community – from enterprise teams to educators and nonprofits. Its easy-to-use AI features are woven into every part of Canva, enabling anyone to spark inspiration, streamline workflows, and scale their content. In fact, our AI tools have been used more than 10 billion times to date.*

**- Canva Press Release, 10/24**



Note: We assume zero users in the launch month. Source: Canva announcements (10/23, 10/24, 5/25)

# Enterprise SaaS Incumbent AI Rollouts = Canva Magic Studio – 10/23

## Canva Magic Studio – 10/23-5/25, per Canva

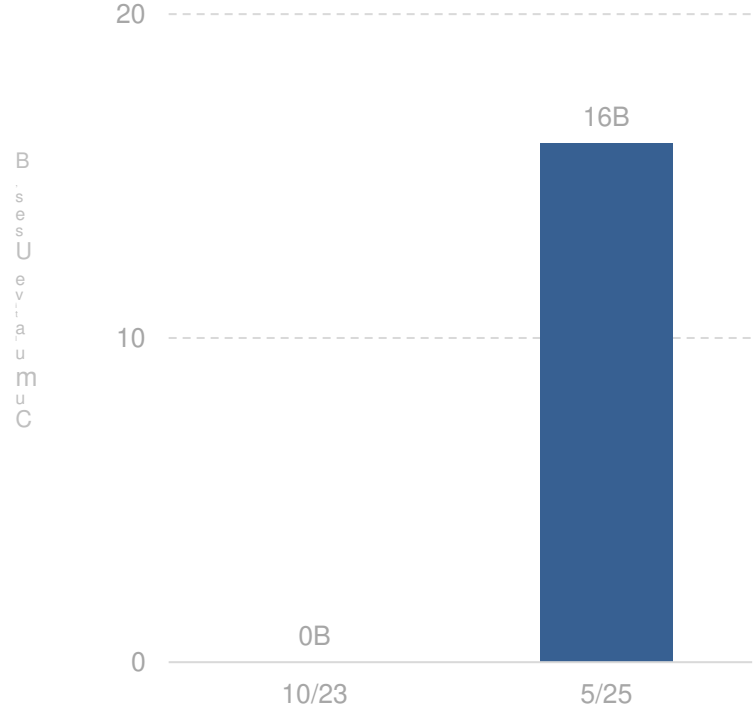


*With Magic Studio there's no need to toggle between multiple AI tools or learn lots of different software – all the best of AI is at your fingertips. Created for the 99% of the world without complex design skills, Magic Studio is jam-packed with easy-to-use AI-powered features across every part of Canva to help you work smarter.*

**- Canva Press Release, 10/23**

*Magic Studio is designed to supercharge creativity across our entire community – from enterprise teams to educators and nonprofits. Its easy-to-use AI features are woven into every part of Canva, enabling anyone to spark inspiration, streamline workflows, and scale their content. In fact, our AI tools have been used more than 10 billion times to date.*

**- Canva Press Release, 10/24**



注意：我们假设发布月份的用户数为零。来源：Canva 公告 (10/23, 10/24, 5/25)



Enterprise SaaS Incumbent AI Rollouts = Salesforce Agentforce – 9/24

Enterprise SaaS Incumbent AI Rollouts = Salesforce Agentforce – 9/24

Salesforce Agentforce – 12/24-2/25, per Salesforce

Salesforce Agentforce – 12/24-2/25, per Salesforce



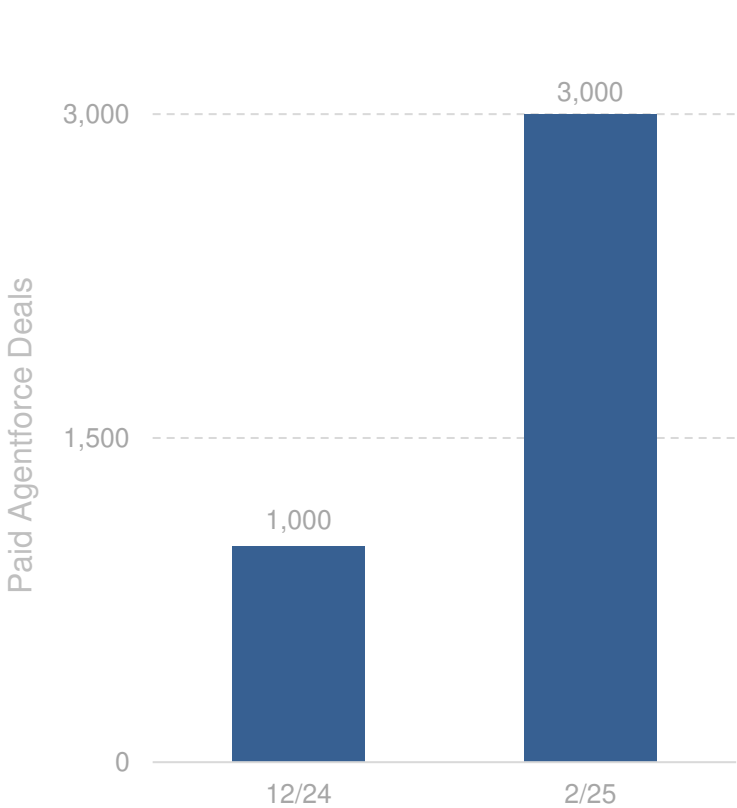
We ended this year with \$900MM in Data Cloud and AI ARR. It grew 120% year over year. We've never seen products grow at these levels, especially Agentforce...

...Just 90 days after it went live, we've already had 3,000 paying Agentforce customers who are experiencing unprecedented levels of productivity, efficiency, and cost savings...

...Data Cloud is the fuel that powers Agentforce and our customers are investing in it. And Data Cloud surpassed 50 trillion, that's trillion with a T, records, doubling year over year as customers increase their consumption and investment in our data platform.

- Salesforce Co-Founder & CEO Mark Benioff, 2/25

Number of Paid Agentforce Deals Signed



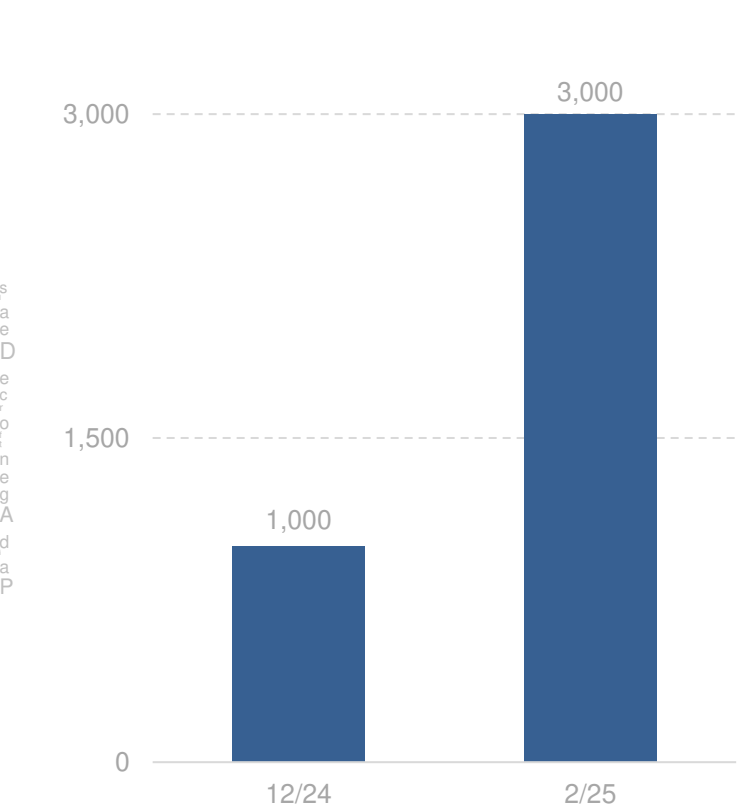
今年年底，我们在 Data Cloud 和 AI 方面的 ARR 达到 9 亿美元。同比增长 120%。我们从未见过产品以这种水平增长，尤其是 Agentforce...

… 在线仅仅 90 天后，我们已经拥有 3,000 家付费 Agentforce 客户，他们正在体验前所未有的生产力、效率和成本节约水平 …

…Data Cloud 是 Agentforce 的动力，我们的客户正在对其进行投资。Data Cloud 超过了 50 万亿条记录，即万亿级别的记录，同比增长了一倍，因为客户增加了他们在我们数据平台上的消费和投资。

- Salesforce 联合创始人兼首席执行官 Mark Benioff, 2/25

已签署的 Agentforce 付费交易数量



Note: Agentforce was announced on 9/12/24 but became generally available on 10/29/24. We assume zero users in the launch month.  
Source: Salesforce announcements (10/24, 12/24, 2/25)

注意：Agentforce 于 2024 年 9 月 12 日发布，但于 2024 年 10 月 29 日全面上市。我们假设发布月份的用户数为零。来源：Salesforce 公告（10/24、12/24、2/25）

Horizontal Enterprise Platform =

SaaS Incumbents

Or

**Large Language Model Challengers?**

横向企业平台 =

SaaS Incumbents

Or

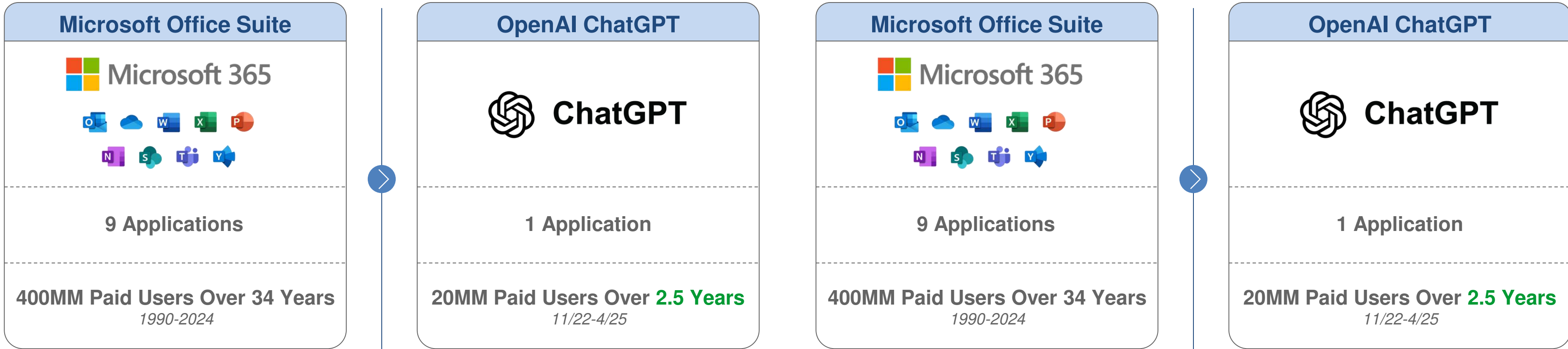
**Large Language Model Challengers?**

OpenAI ChatGPT =  
Potential Horizontal Enterprise Platform?...

OpenAI ChatGPT =  
Potential Horizontal Enterprise Platform?...

OpenAI = Next-Gen All-in-One Enterprise Platform?

OpenAI = 下一代一体化企业平台?



Source: Microsoft (1/24), Office365 Pros, OpenAI, The Information (4/25) (link)

Source: Microsoft (1/24), Office365 Pros, OpenAI, The Information (4/25) (link)



# ...OpenAI ChatGPT = Potential Horizontal Enterprise Platform?

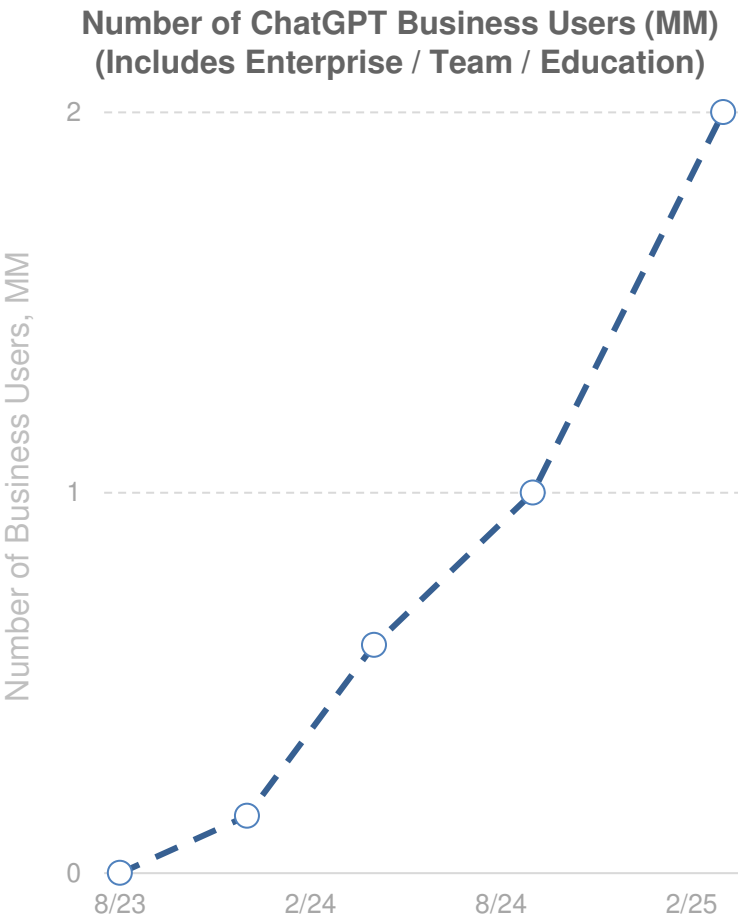
ChatGPT Enterprise – 8/23-3/25, per OpenAI, *Bloomberg*, *Reuters*, & *The Wall Street Journal*



Since ChatGPT’s launch just nine months ago, we’ve seen teams adopt it in over 80% of Fortune 500 companies. We’ve heard from business leaders that they’d like a simple and safe way of deploying it in their organization. Early users of ChatGPT Enterprise...are redefining how they operate and are using ChatGPT to craft clearer communications, accelerate coding tasks, rapidly explore answers to complex business questions, assist with creative work, and much more.

ChatGPT Enterprise removes all usage caps and performs up to two times faster [vs. ChatGPT Free]...  
...ChatGPT Enterprise also provides unlimited access to advanced data analysis, previously known as Code Interpreter.

- ChatGPT Enterprise Release Statement, 8/23



Note: We assume zero users in the launch month. Source: OpenAI announcements (12/23, 4/24, 9/24, 3/25), Bloomberg (4/24), Reuters (9/24), The Wall Street Journal (3/25)

# ...OpenAI ChatGPT = Potential Horizontal Enterprise Platform?

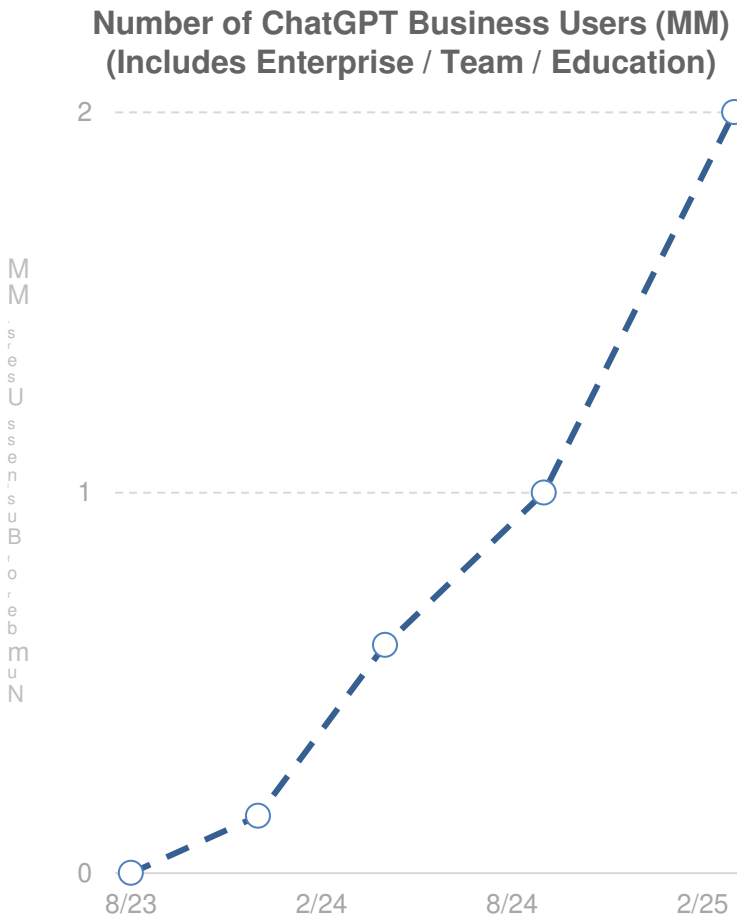
ChatGPT Enterprise – 8/23-3/25, 根据 OpenAI、彭博社、路透社和《华尔街日报》报道



自从 ChatGPT 推出仅九个月以来，我们已经看到超过 80% 的财富 500 强公司中的团队采用了它。我们从商业领袖那里听说，他们希望有一种简单而安全的方式在他们的组织中部署它。ChatGPT Enterprise 的早期用户 ..... 正在重新定义他们的运营方式，并且正在使用 ChatGPT 来制作更清晰的沟通，加速编码任务，快速探索复杂业务问题的答案，协助创意工作等等。

ChatGPT Enterprise 取消了所有使用上限，并且执行速度提高了两倍 [与 ChatGPT Free 相比]……ChatGPT Enterprise 还提供对高级数据分析的无限制访问，以前称为 Code Interpreter。

- ChatGPT Enterprise 发布声明，8/23



Note: We assume zero users in the launch month. 来源：OpenAI 公告（12/23、4/24、9/24、3/25）、彭博社（4/24）、路透社（9/24）、华尔街日报（3/25）

AI Monetization Possibilities – Enterprise =

Horizontal Platform  
& / Or  
**Specialized Software?**

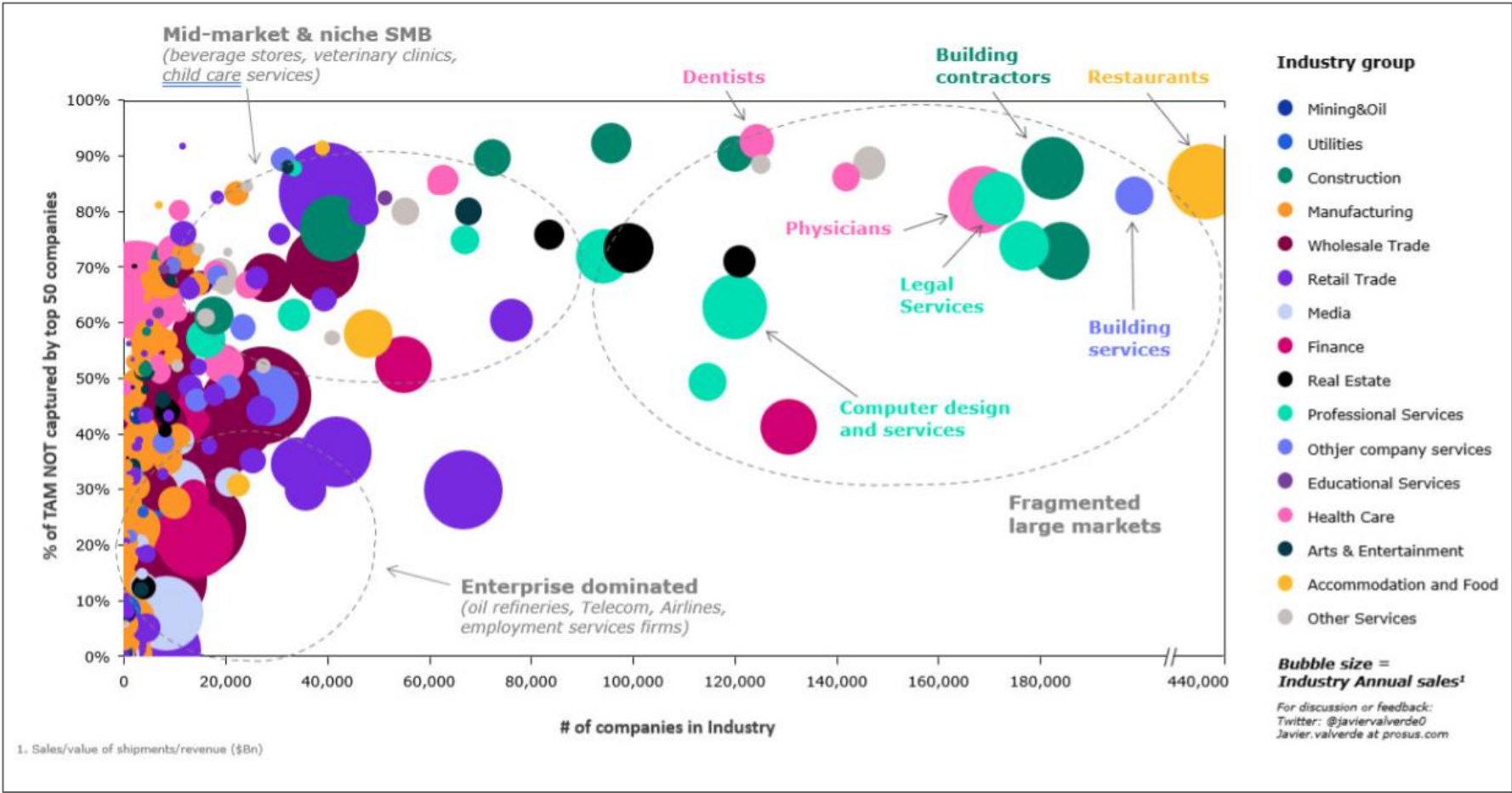
AI 货币化可能性 – 企业 =

横向平台 & / 或专  
业软件?

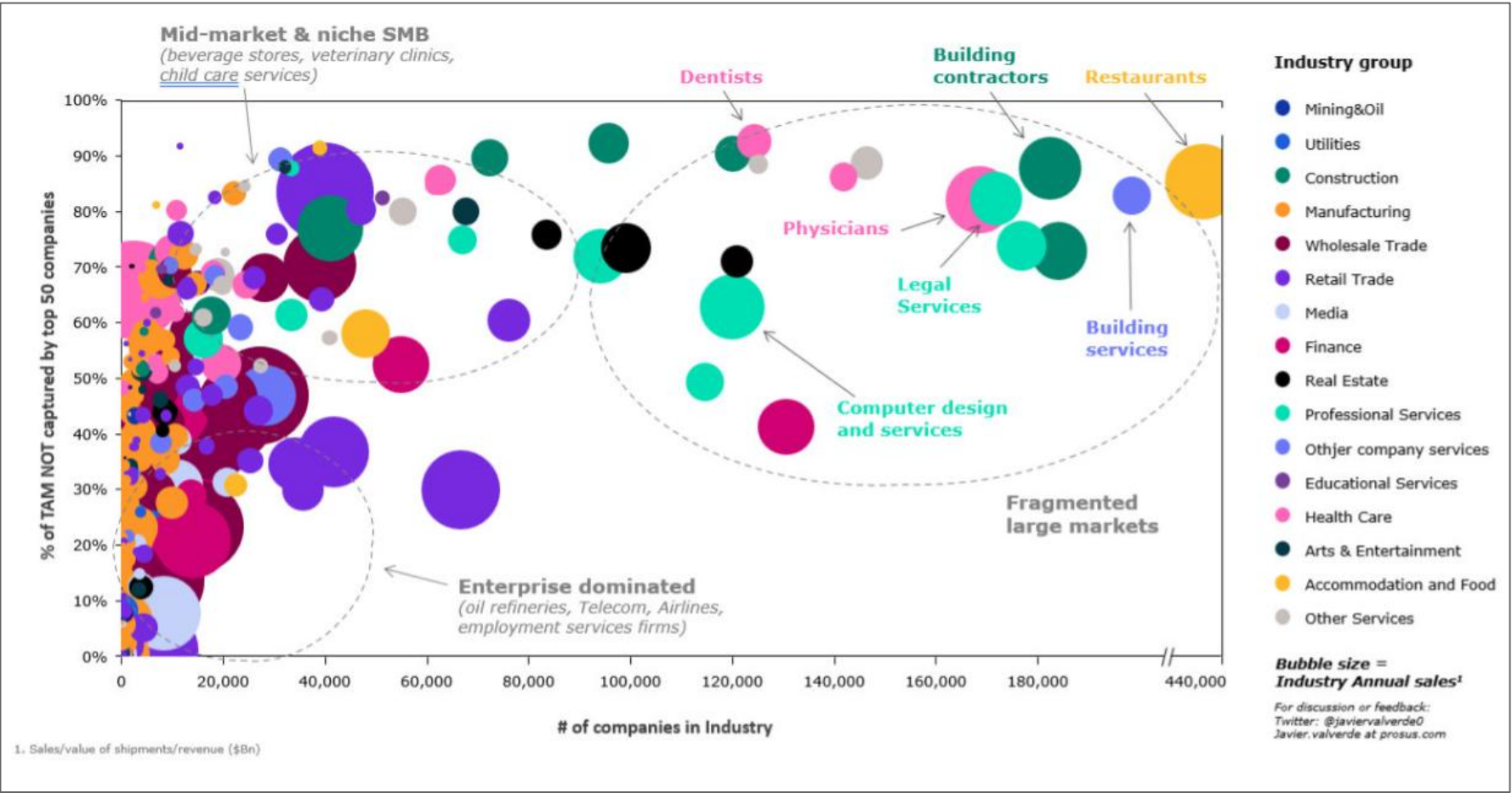
AI Monetization – Enterprise =  
Specialized Software Opportunities in Fragmented Markets, per Prosus

AI 货币化 – 企业 = 在分散市场中的专业软件机会，根据 Prosus 的数据

USA Industries by Number of Companies & Market Share – 2024, per Prosus



美国各行业公司数量和市场份额 – 2024 年，根据 Prosus 的数据



AI-Enabled Specialized Software @  
Large Service Industries =

Growing Very Quickly...

Software Engineering  
Product Development  
Healthcare  
Legal  
Customer Service  
Financial Services

AI 赋能的专业软件 @ 大型服务业 =

快速增长 .....

软件工程 产品开发 医  
疗保健 法律 客户服务  
金融服务

*AI-Enabled Specialized Software Companies @  
Large Service Industries =*

*Growing Very Quickly...*

***Software Engineering***

大型服务行业的 AI 赋能的专业软件公司 =

快速增长 .....

软件工程

# Specialized AI – Software Engineering (Code Editor) = Anysphere Cursor AI ARR @ \$1MM to \$300MM in Twenty-Five Months

# 专用 AI – 软件工程（代码编辑器） =Anysphere Cursor AI ARR @ 100 万美元至 3 亿美元，历时 25 个月

## Anysphere Cursor AI – 3/23-4/25, per Anysphere

## Anysphere Cursor AI –3/23-4/25, 根据 Anysphere



*Something beautiful is happening to code...our aim with Cursor is to continue to lead this shift, by building a magical tool that will one day write all the world's software...*

*...Already, in Cursor, hours of hunting for the right primitives are being replaced by instant answers. Mechanical refactors are being reduced to single 'tabs.' Terse directives are getting expanded into working source. And thousand-line changes are rippling to life in seconds.*

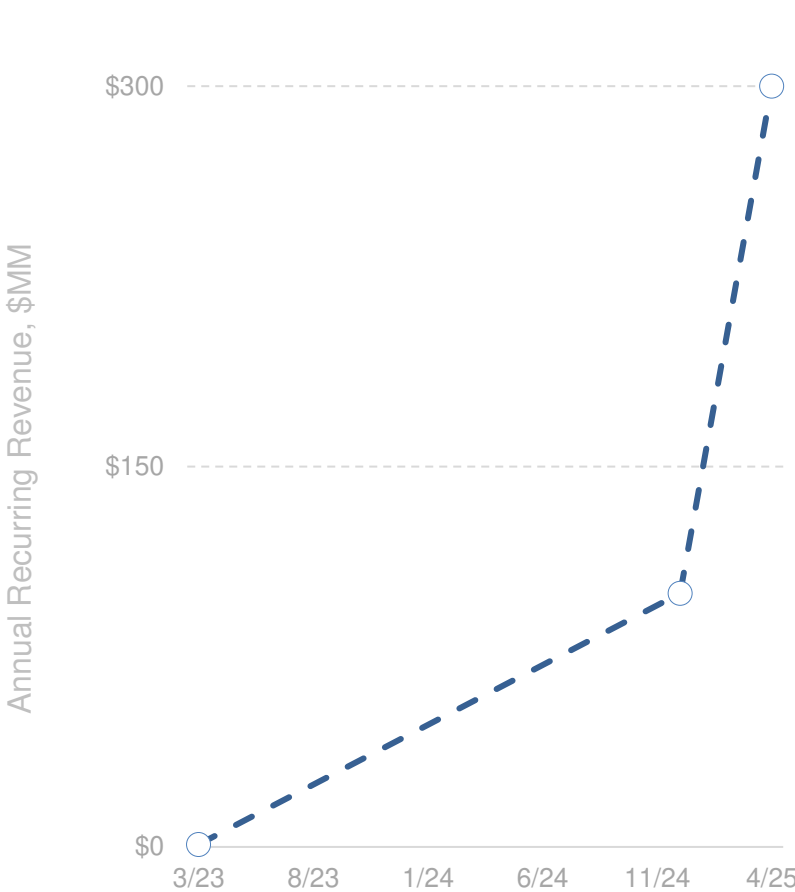
**- Anysphere Press Release (8/24)**

*...We're delighted to report that Cursor is now used by millions of programmers as their editor of choice. Our proprietary models now generate more code than almost any LLMs in the world and edit over a billion characters per day. Our business is large and fast growing, having exceeded \$100MM in recurring revenue.*

**- Anysphere Team (8/24 & 1/25)**

Note: Cursor launched in 4/23. We show 3/23 as the first datapoint with an assumed \$0 in ARR. Source: Cursor / Anysphere (8/24, 11/24 & 1/25), Anysphere Co-Founder & CEO Michael Truell via Lenny's Newsletter, 'The rise of Cursor: The \$300M ARR AI tool that engineers can't stop using' (5/1/25)

Annual Recurring Revenue (ARR) (\$MM)



代码正在发生一些美好的事情 ..... 我们使用 Cursor 的目标是继续引领这一转变，通过构建一个神奇的工具，有一天将编写世界上所有的软件 .....

..... 在 Cursor 中，寻找正确原语的数小时时间已经被即时答案所取代。机械重构正在减少到单个“选项卡”。简洁的指令正在扩展到工作源。成千上万行的更改在几秒钟内就变得栩栩如生。

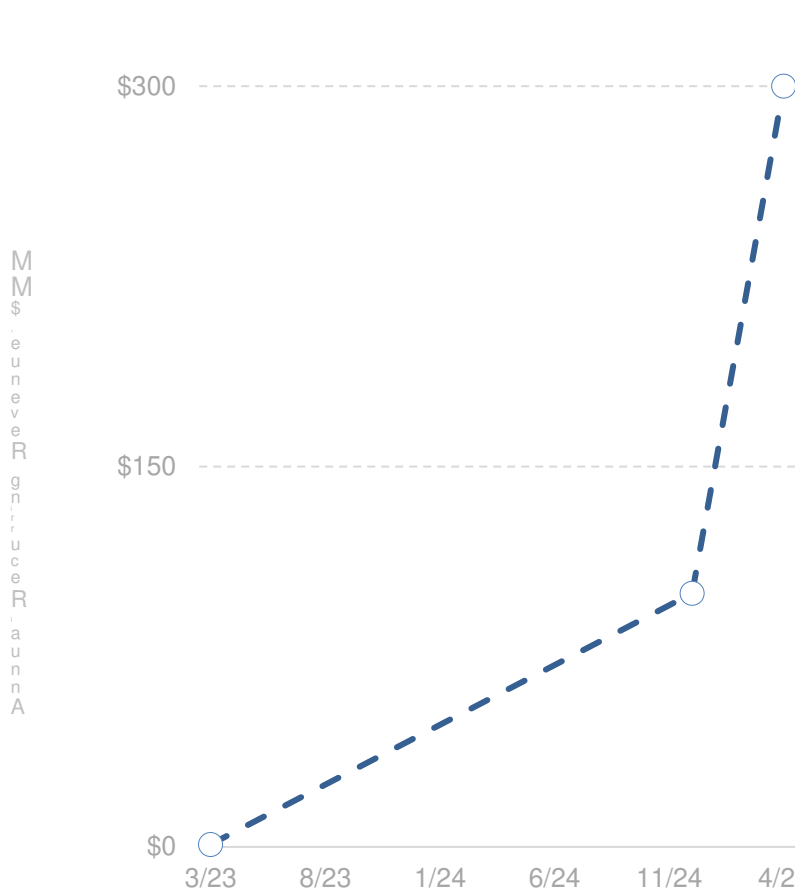
**- Anysphere 新闻稿 (8/24)**

..... 我们很高兴地报告，数百万程序员现在正在使用 Cursor 作为他们选择的编辑器。我们专有的模型现在生成的代码比世界上几乎任何 LLM 都多，并且每天编辑超过 10 亿个字符。我们的业务规模庞大且快速增长，经常性收入已超过 1 亿美元。

**- Anysphere Team (8/24 & 1/25)**

注意：Cursor 于 4/23 推出。我们将 3/23 显示为第一个数据点，并假设 ARR 为 0 美元。来源：Cursor / Anysphere（8/24、11/24 和 1/25），Anysphere 联合创始人兼首席执行官 Michael Truell via Lenny's Newsletter, 'Cursor 的崛起：工程师们无法停止使用的 3 亿美元 ARR AI 工具'（5/1/25）

年度经常性收入 (ARR) (\$MM)



*AI-Enabled Specialized Software Companies @  
Large Service Industries =*

*Growing Very Quickly...*

***Product Development***

AI 赋能的专业软件公司 @ 大型服务行业 =

快速增长 .....

产品开发

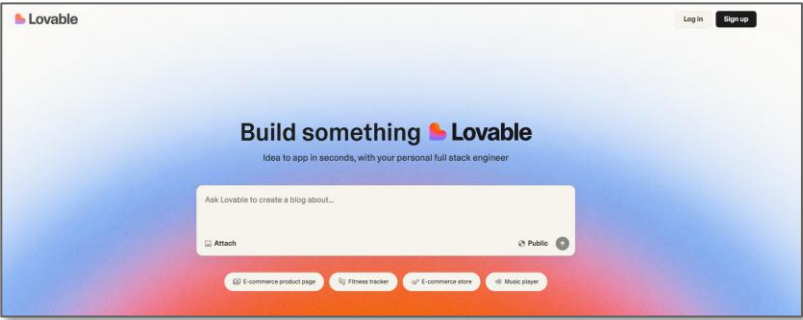


# Specialized AI – Product Development (No-Code Product-Building) = Lovable ARR +13x to \$50MM in Five Months

# 专业人工智能 – 产品开发（无代码产品构建）=Lovable ARR +13x 在五个月内达到 5000 万美元

Lovable – 12/24-5/25

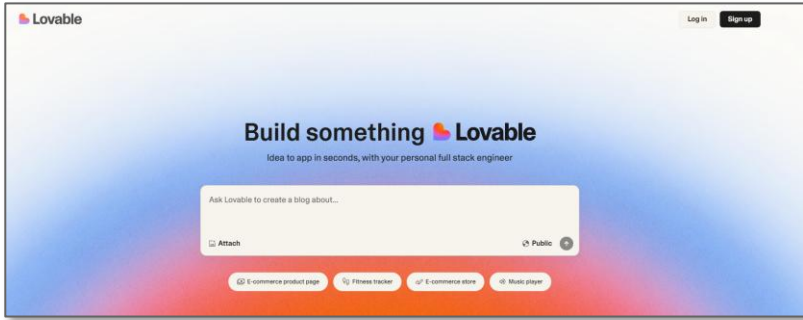
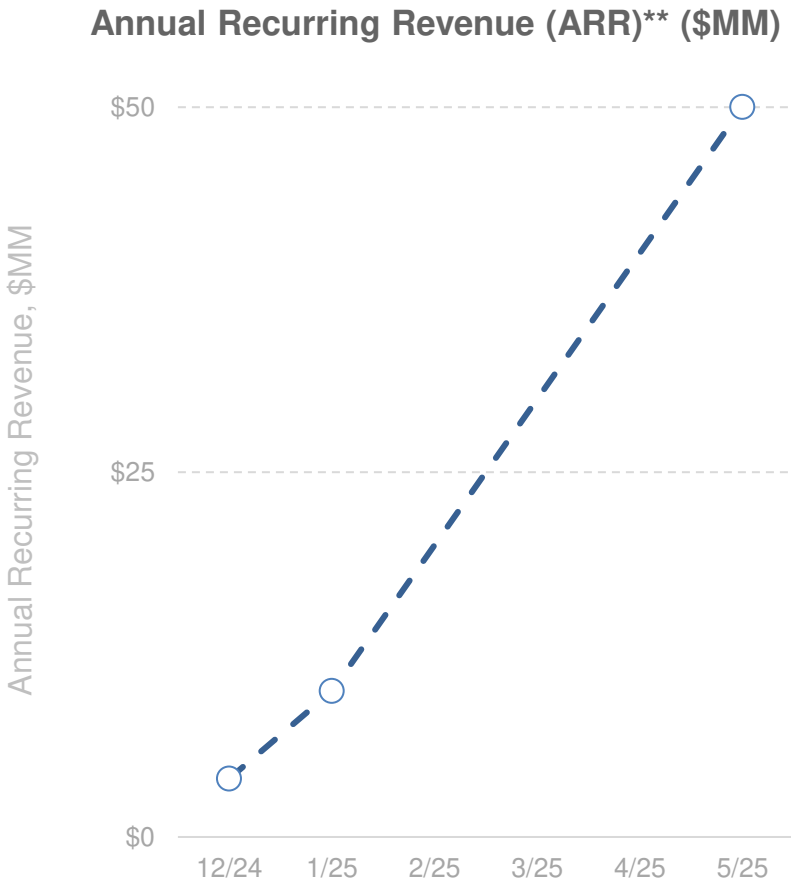
令人喜爱 – 12/24-5/25



Note: Lovable is an AI-powered application development platform that enables users to create full-stack web applications by describing their ideas in natural language. The platform translates these descriptions into functional applications, handling frontend and backend code generation, database integration, and deployment.

*The opportunity here is immense. We are on the verge of a paradigm shift where the barriers to building software-based products disappear. Now, anyone can become an entrepreneur, launch a product and build a business in minutes.*

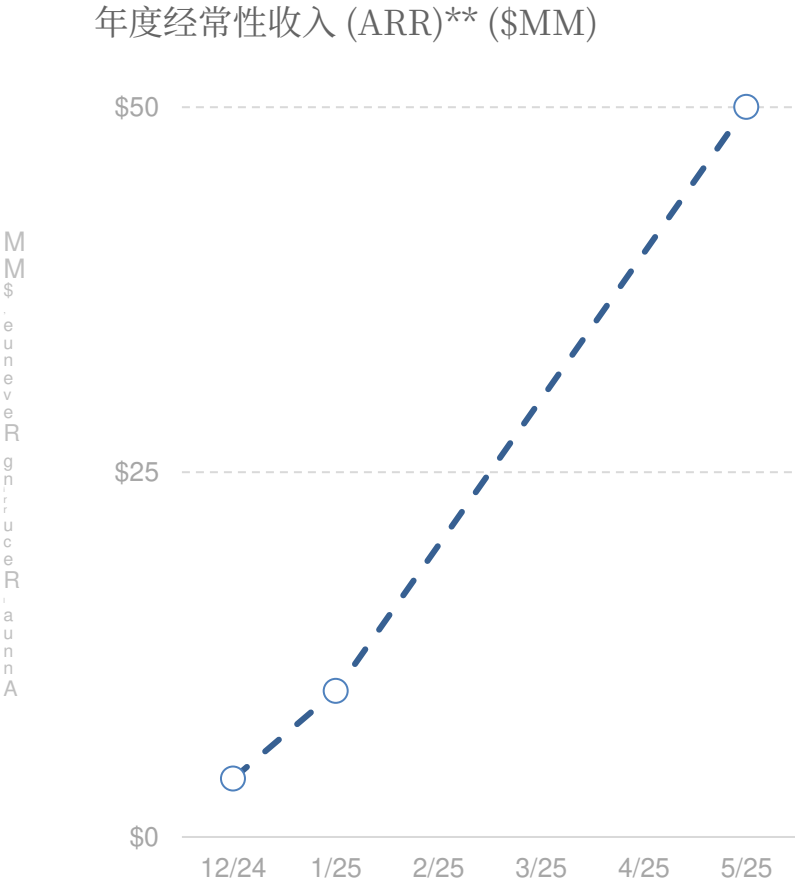
**- Frederik Cassel, Creandum, ‘Backing Lovable: Move Fast and Make Things,’ 2/25\***



注意：Lovable 是一个 AI 驱动的应用程序开发平台，使用户能够通过以自然语言描述他们的想法来创建全栈 Web 应用程序。该平台将这些描述转换为功能性应用程序，处理前端和后端代码生成、数据库集成和部署。

这里的机会有巨大的。我们正处在一个范式转变的边缘，构建基于软件的产品障碍正在消失。现在，任何人都可以成为企业家，在几分钟内推出产品并建立业务。

**- Frederik Cassel, Creandum,‘支持 Lovable：快速行动并创造价值’， 2/25\***





*AI-Enabled Specialized Software Companies @  
Large Service Industries =*

*Growing Very Quickly...*

***Healthcare***

AI 赋能的专业软件公司 @ 大型服务业 =

快速增长 .....

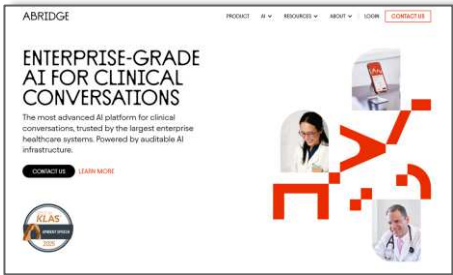
医疗保健

# Specialized AI – Healthcare (Clinical Conversations) = Abridge @ \$50MM to \$117MM CARR in ~Five Months

# 专业人工智能 – 医疗保健（临床对话） =Abridge @ 5000 万美元至 1.17 亿美元 CARR，耗时 ~ 五个月

Abridge – 10/24-3/25, per Abridge & The Information

Abridge – 10/24-3/25, 根据 Abridge & The Information



Yazdi Bagli, Kaiser’s EVP of IT and enterprise business services, said he believes [Kaiser Permanente’s] Abridge partnership is one of the largest generative AI deployments in health care...

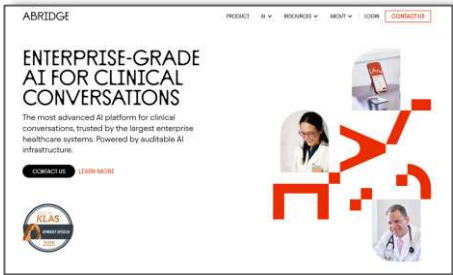
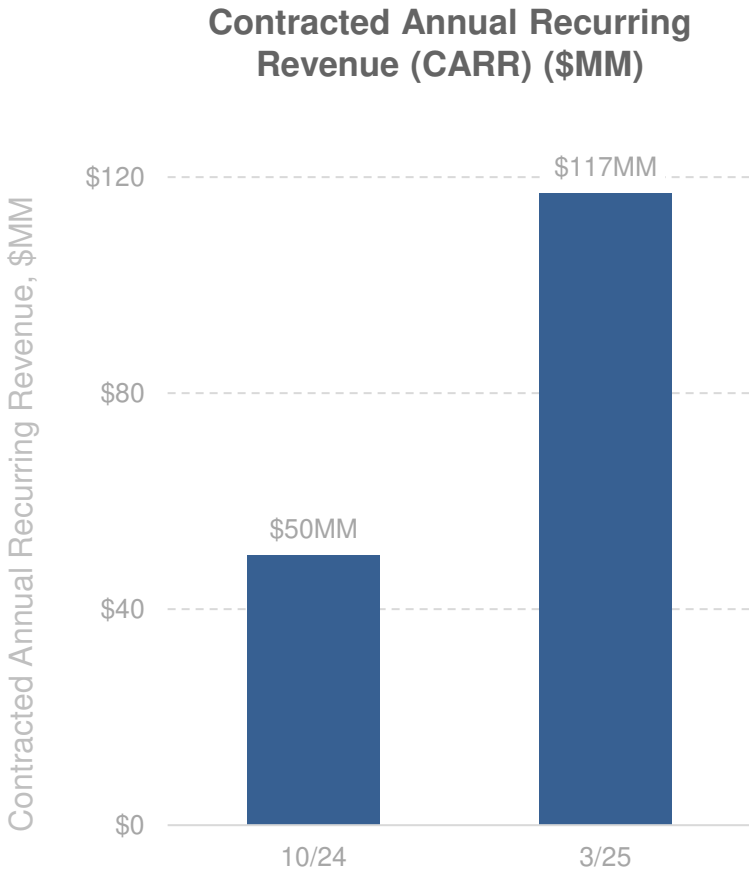
...The national rollout includes more than 25,000 doctors and clinicians, 40 hospitals, and north of 600 medical offices...

...The feedback from doctors has been effusive: ‘It saved my marriage.’ And: ‘You’d have to take it away from my cold, dying hands.’

- Fortune Magazine (2/25)

We are incredibly proud of our partnership with Kaiser – where a majority of Kaiser doctors are using Abridge to summarize patient visits, with over 10 million completed to date. As one of our earliest deployments, it is a great example of how we are building alongside our many hospital partners and helping them grow with Abridge.

- Abridge CFO Sagar Sanghvi (5/25)



凯撒信息技术和企业业务服务执行副总裁 Yazdi Bagli 表示，他认为 [Kaiser Permanente 的] Abridge 合作关系是医疗保健领域最大的生成式人工智能部署之一……

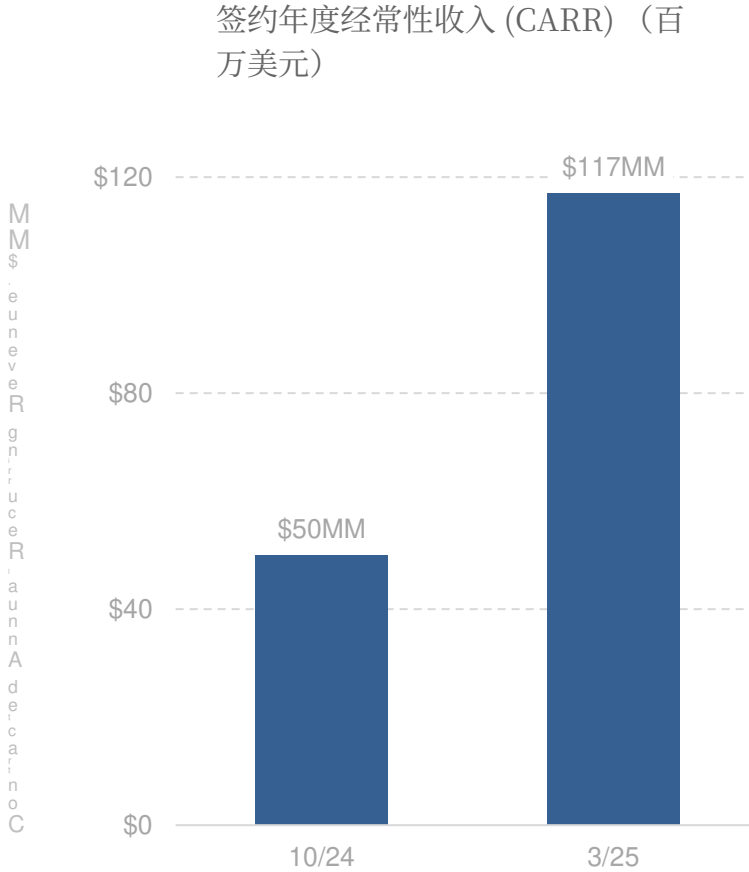
…… 全国推广包括超过 25,000 名医生和临床医生、40 家医院和 600 多家医疗办公室……

…… 来自医生的反馈非常热情：“它拯救了我的婚姻。” 还有：“你得从我冰冷、垂死的手中夺走它。”

- 《财富》杂志（2 月 25 日）

我们非常自豪能与 Kaiser 合作 –，Kaiser 的大多数医生都在使用 Abridge 来总结患者就诊情况，迄今已完成超过 1000 万次。作为我们最早的部署之一，这是一个很好的例子，说明我们如何与众多医院合作伙伴并肩建设，并帮助他们与 Abridge 一起成长。

- Abridge 首席财务官 Sagar Sanghvi (5/25)



Note: 3/25 figure is quoted as being as of Q1:25. We conservatively assume this maps to 3/25. Abridge’s CARR goes live within weeks of contracting. Source: Abridge (12/24 & 5/25), Fortune (2/25), The Information (10/24 & 5/25) ([link](#) & [link](#))

Note: 3/25 figure is quoted as being as of Q1:25. We conservatively assume this maps to 3/25. Abridge’s CARR goes live within weeks of contracting. Source: Abridge (12/24 & 5/25), Fortune (2/25), The Information (10/24 & 5/25) ([link](#) & [link](#))

*AI-Enabled Specialized Software Companies @  
Large Service Industries =*

*Growing Very Quickly...*

***Legal***

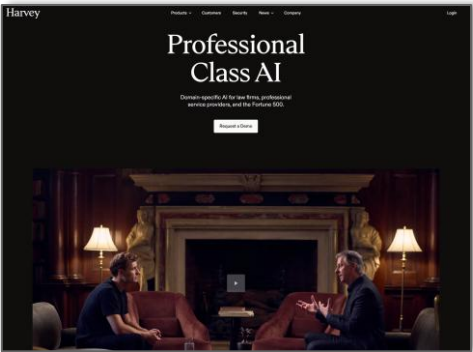
在大型服务行业中支持 AI 的专用软件公司 =

快速增长 ……

法律

# Specialized AI – Legal (Workflows) = Harvey @ \$10MM to \$70MM ARR in Fifteen Months, per *The Information & Business Insider*

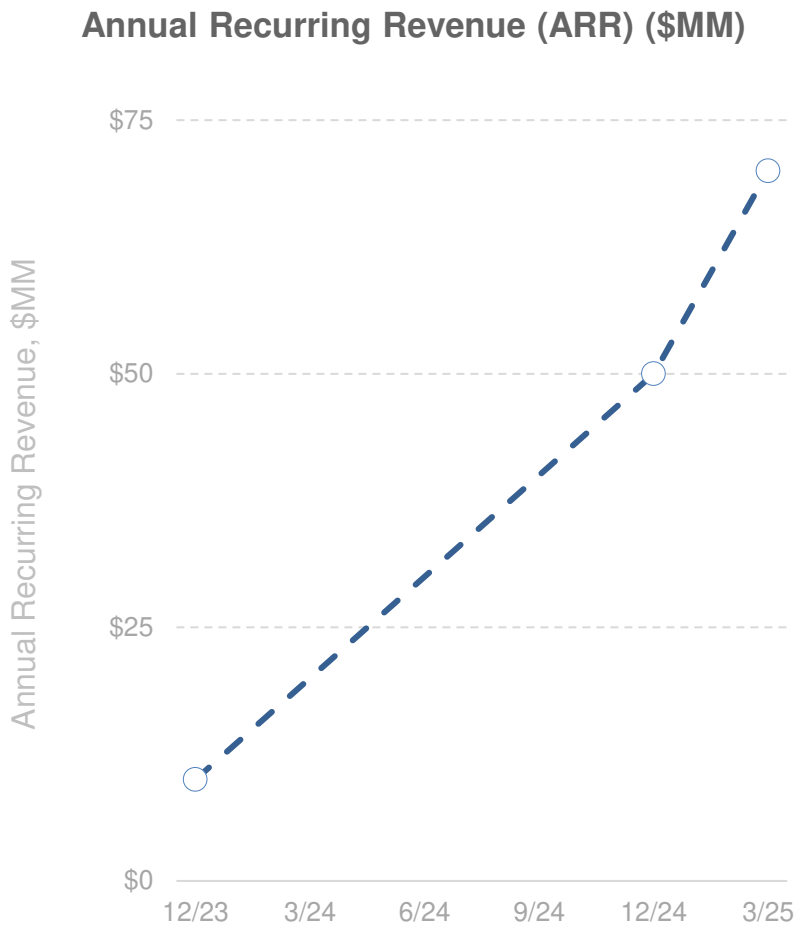
Harvey – 12/23-4/25, per *The Information & Business Insider*



*In 2024, we saw 4x annual recurring revenue (ARR) growth and expanded from 40 customers to 235 customers in 42 countries, including the majority of the top 10 USA law firms.*

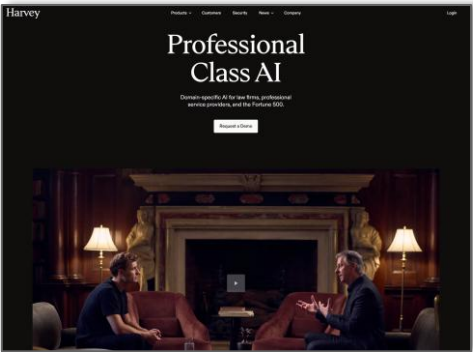
*We’ve also seen the legal and professional services industry shift faster than ever before. Lawyers are adopting technology at an unprecedented rate, centuries-old firms are experimenting with new business models, and enterprises are driving significant savings with AI-enabled workflows. The pace of change will only accelerate in 2025.*

**- Harvey Co-Founder & CEO Winston Weinberg & Co-Founder & President Gabe Pereyra (2/25)**



# Specialized AI – Legal (Workflows) = Harvey @ \$10MM to \$70MM ARR in Fifteen Months, per *The Information & Business Insider*

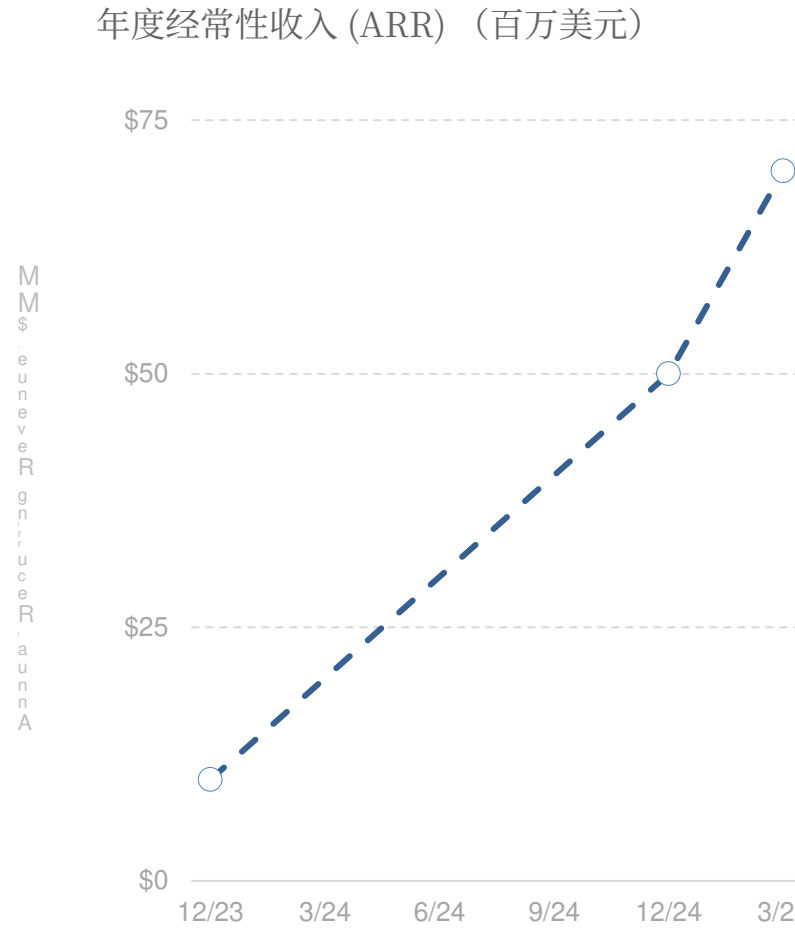
Harvey –12/23-4/25, 根据 The Information & Business Insider



2024 年，我们实现了 4 倍的年度经常性收入 (ARR) 增长，客户数量从 40 家扩展到 42 个国家 / 地区的 235 家，其中包括美国排名前 10 的大多数律师事务所。

*We’ve also seen the legal and professional services industry 变革速度比以往任何时候都快。律师们正以空前的速度采用技术，拥有数百年历史的公司正在尝试新的商业模式，企业正在通过支持人工智能的工作流程大幅节省成本。2025 年，变革的步伐只会加快。*

**- Harvey 联合创始人兼首席执行官 Winston Weinberg 和联合创始人兼总裁 Gabe Pereyra (2/25)**



Source: Harvey (2/25), The Information estimates (1/25) ([link](#), [link](#)), & Business Insider (5/25) ([link](#))

Source: Harvey (2/25), The Information estimates (1/25) ([link](#), [link](#)), & Business Insider (5/25) ([link](#))

*AI-Enabled Specialized Software Companies @  
Large Service Industries =*

*Growing Very Quickly...*

***Customer Service***

AI 支持的专业软件公司 @ 大型服务行业 =

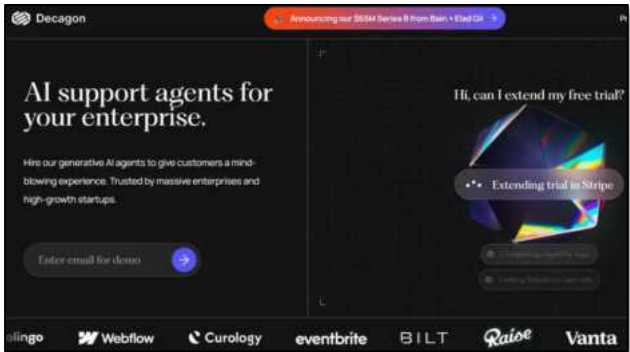
增长非常迅速 ……

客户服务

# Specialized AI – Customer Service (AI Support Agents) = Decagon @ ~\$1MM to \$10MM ARR in One Year

# 专业人工智能 – 客户服务（ AI 支持代理） =Decagon @ ~ 一年内达到 100 万美元至 1000 万美元的 ARR

Decagon – 2023-2024, per Decagon

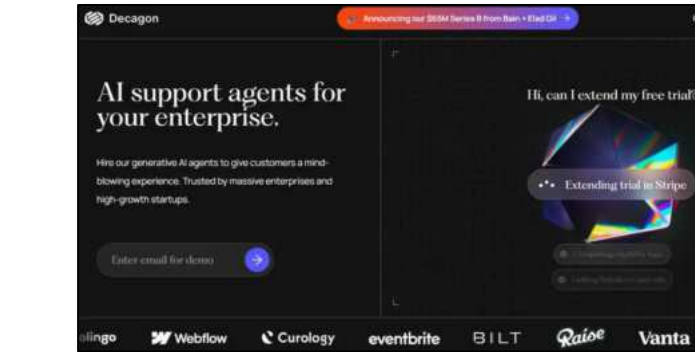
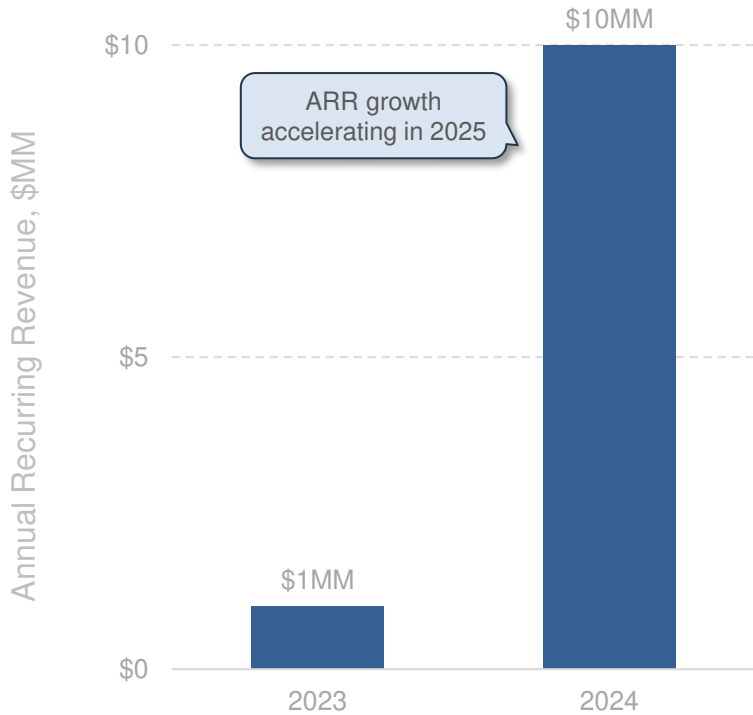


AI is often seen as destroying jobs, but at Decagon, we believe the opposite. Our AI agents are enhancing jobs, not replacing them...

...In a few years, every company will have AI agents running their customer experiences. Customer support staff are no longer fielding routine tasks; they are now becoming AI managers – configuring, training and overseeing the AI agents that handle repetitive work.

- Decagon Co-Founder & CEO Jesse Zhang (10/24)

Annual Recurring Revenue (ARR) (\$MM)



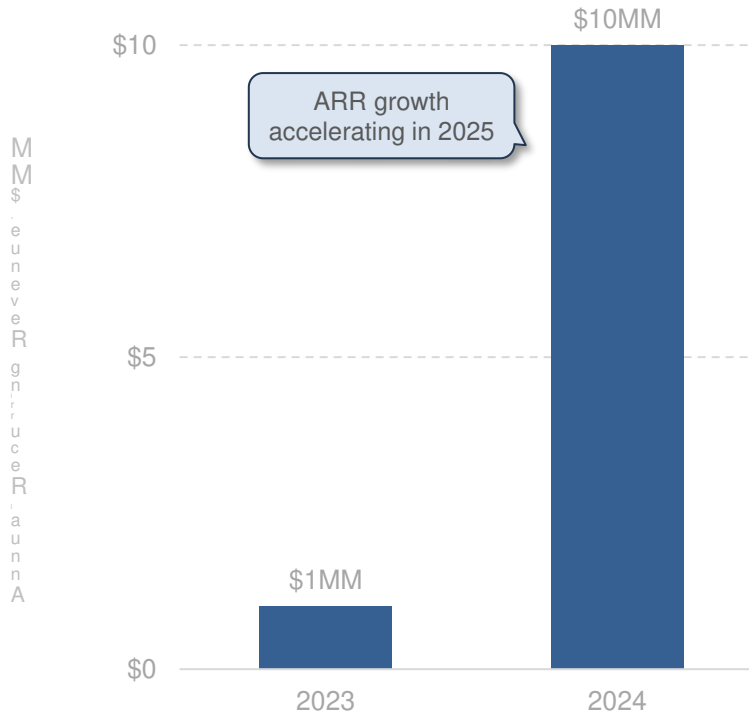
人们通常认为人工智能会摧毁就业机会，但在 Decagon，我们相信事实恰恰相反。我们的人工智能代理正在增强就业机会，而不是取代它们 ...

... 在几年内，每家公司都将运行人工智能代理来改善他们的客户体验。客户支持人员不再处理日常任务；他们现在正在成为人工智能经理 – 配置、培训和监督处理重复性工作的人工智能代理。

- Decagon 联合创始人兼首席执行官 Jesse Zhang (10/24)

Decagon – 2023-2024, per Decagon

年度经常性收入（ARR）（百万美元）



Note: Source: Decagon (12/23, 10/24, 12/24)

注意：来源：Decagon (12/23, 10/24, 12/24)

*AI-Enabled Specialized Software Companies @  
Large Service Industries =*

*Growing Very Quickly...*

***Financial Services***

AI 赋能的专业软件公司 @ 大型服务业 =

快速增长 .....

金融服务

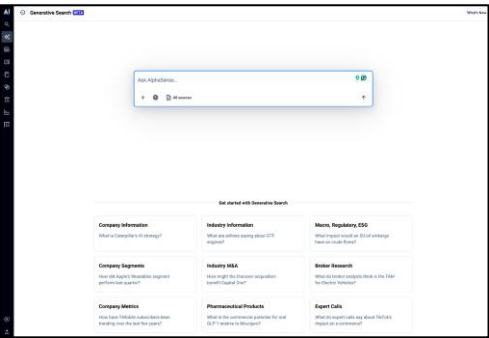
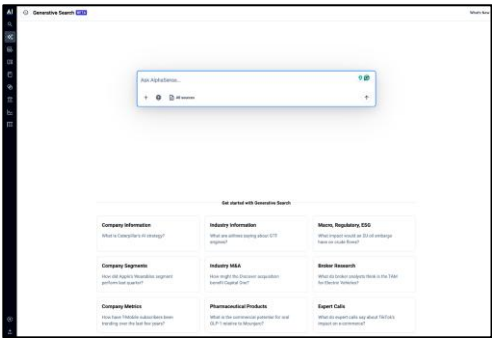


# Specialized AI – Financial Services (Research & Analysis) = AlphaSense @ ~\$150MM to ~\$420MM in Two Years

# 专业人工智能 – 金融服务（研究与分析） =AlphaSense @ ~\$150MM 至 ~\$420MM 两年内

AlphaSense – 2022-2024, per AlphaSense

AlphaSense – 2022-2024, 根据 AlphaSense



*We are at a tipping point where AI-driven insights are no longer a luxury but a necessity – every company’s market value is the sum of the decisions it makes.*

*Surpassing \$400 million in ARR and our rapid growth are clear signals that businesses are recognizing the transformative power of our end-to-end market intelligence platform.*

*As we scale, our focus remains on product and technology innovation, ensuring we deliver high-value solutions and cutting-edge AI and smart workflow capabilities to our customers.*

**- AlphaSense Co-Founder & CEO Jack Kokko (3/25)**

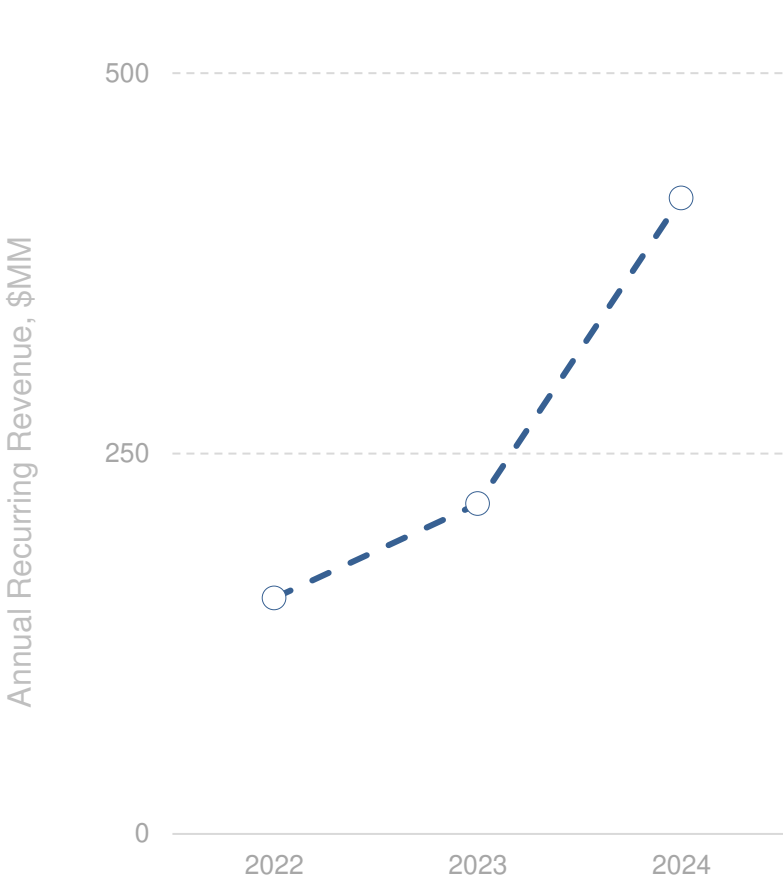
我们正处于一个临界点，人工智能驱动的洞察不再是一种奢侈品，而是一种必需品 – 每个公司的市场价值是其所做决定的总和。

超过 4 亿美元的 ARR 和我们的快速增长清楚地表明，企业正在认识到我们端到端市场情报平台的变革力量。

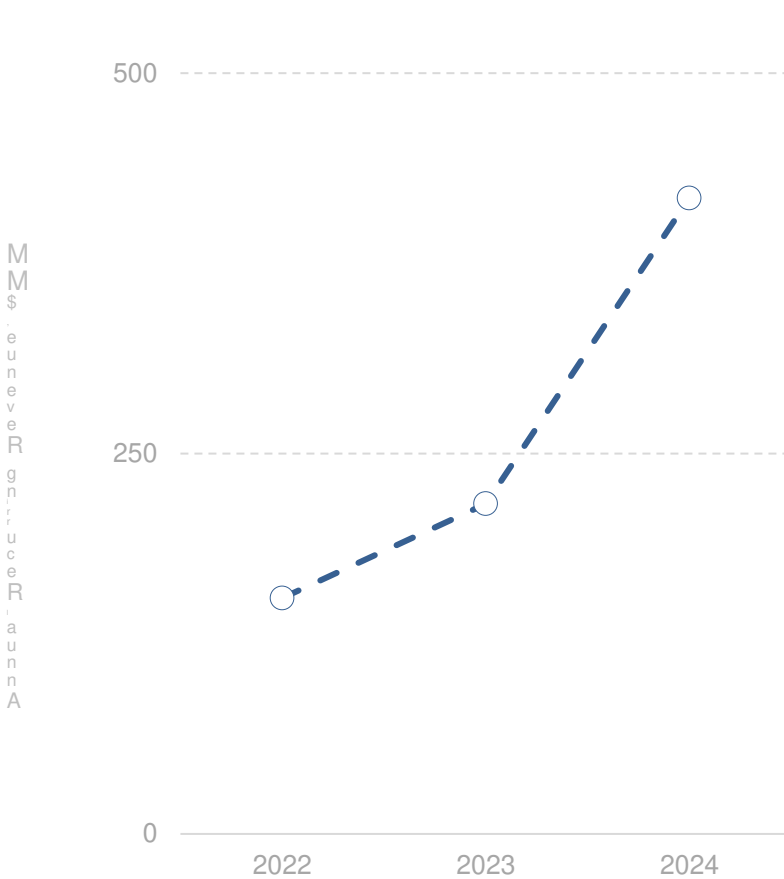
随着我们的扩展，我们的重点仍然是产品和技术创新，确保我们为客户提供高价值的解决方案和尖端的人工智能和智能工作流程功能。

- AlphaSense 联合创始人兼首席执行官 Jack Kokko (3/25)

Annual Recurring Revenue (ARR) (\$MM)



年度经常性收入（ARR）（百万美元）



Source: AlphaSense (3/25)

Source: AlphaSense (3/25)

Next AI Use Case Frontiers =

Broad & Varied

下一个 AI 用例前沿 =

Broad & Varied

# Next AI Use Case Frontiers = Broad & Varied

# 下一个人工智能用例前沿 = 广泛且多样

## Next AI Use Case Frontiers – 5/25

## 下一个人工智能用例前沿 – 5/25

Highlights = Pages 246-247

<b>Medical Discovery &amp; Development</b> 	<b>Precision Manufacturing</b> 	<b>Multi-Purpose Robotics</b> 	<b>Autonomous Scientific Research</b> 	<b>Supply Chain Optimization</b> 
<b>Cybersecurity &amp; Threat Detection</b> 	<b>Personalized Education</b> 	<b>Autonomous Finance</b> 	<b>Environmental &amp; Climate Monitoring</b> 	<b>Energy Grid Management</b> 

Highlights = Pages 246-247

<b>Medical Discovery &amp; Development</b> 	<b>Precision Manufacturing</b> 	<b>Multi-Purpose Robotics</b> 	<b>Autonomous Scientific Research</b> 	<b>Supply Chain Optimization</b> 
<b>Cybersecurity &amp; Threat Detection</b> 	<b>Personalized Education</b> 	<b>Autonomous Finance</b> 	<b>Environmental &amp; Climate Monitoring</b> 	<b>Energy Grid Management</b> 

Note: List is not comprehensive. Source: Drug Development & Discovery = Insilico; Precision Manufacturing = Landing AI; Multi-Purpose Robotics = Figure AI; Autonomous Scientific Research = IBM's RoboRXN; Supply Chain Optimization = o9 Solutions; Cybersecurity & Threat Detection = Vectra AI; Personalized Education = Khanmigo; Autonomous Finance = Kasisto; Environmental & Climate Monitoring = ClimateAI; Energy Grid Management = Uplight; BOND analysis

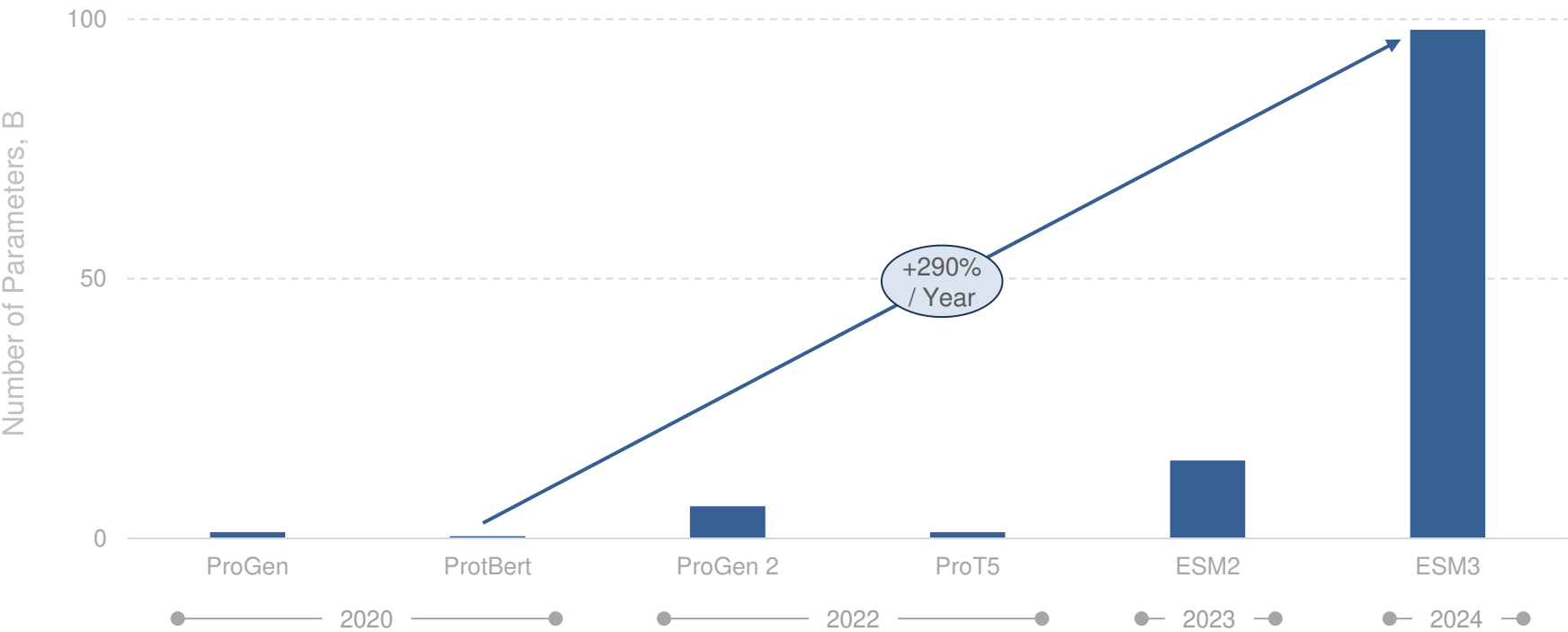
注意：列表不全面。来源：药物开发与发现 = Insilico；精密制造 = Landing AI；多用途机器人 = Figure AI；自主科学研究 = IBM's RoboRXN；供应链优化 = o9 Solutions；网络安全与威胁检测 = Vectra AI；个性化教育 = Khanmigo；自主金融 = Kasisto；环境与气候监测 = ClimateAI；能源网络管理 = Uplight；BOND 分析

# Next AI Use Case Frontier – Protein Sequencing = Model Size +290% Annually to 98 Billion Parameters Over Four Years

# 下一个 AI 用例前沿 – 蛋白质测序 = 模型大小 +290%，四年内每年增长到 980 亿个参数

Size of Major Protein Sequencing Models (B Parameters) – 2020-2024, per Stanford RAISE Health

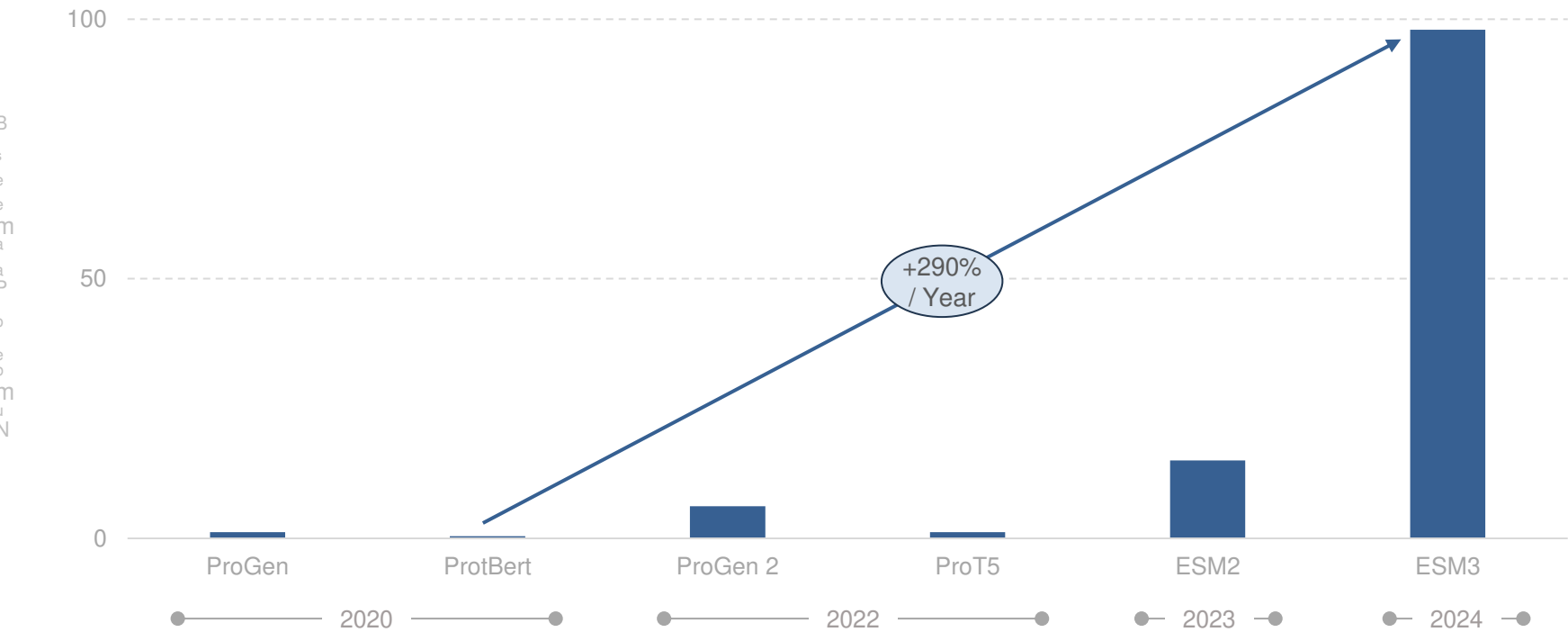
Per Stanford HAI (4/25): *The past year has witnessed remarkable progress in AI models applied to protein sequences. Large-scale machine learning models have improved our ability to predict protein properties, accelerating research in structural biology and molecular engineering...These AI-driven approaches have transformed protein science by minimizing reliance on costly, time-intensive experimental methods, enabling rapid exploration of protein function and design.*



Note: List of models may not be comprehensive.  
Source: Stanford RAISE Health via Nestor Maslej et al., 'The AI Index 2025 Annual Report,' AI Index Steering Committee, Stanford HAI (4/25)

主要蛋白质测序模型的大小（ B 参数） – 2020-2024，数据来源：Stanford RAISE Health

根据 Stanford HAI (4/25) 的说法：过去一年，应用于蛋白质序列的 AI 模型取得了显著进展。大规模机器学习模型提高了我们预测蛋白质属性的能力，加速了结构生物学和分子工程领域的研究 …… 这些 AI 驱动的方法通过最大限度地减少对昂贵的、耗时的实验方法的依赖，从而改变了蛋白质科学，从而能够快速探索蛋白质功能和设计。

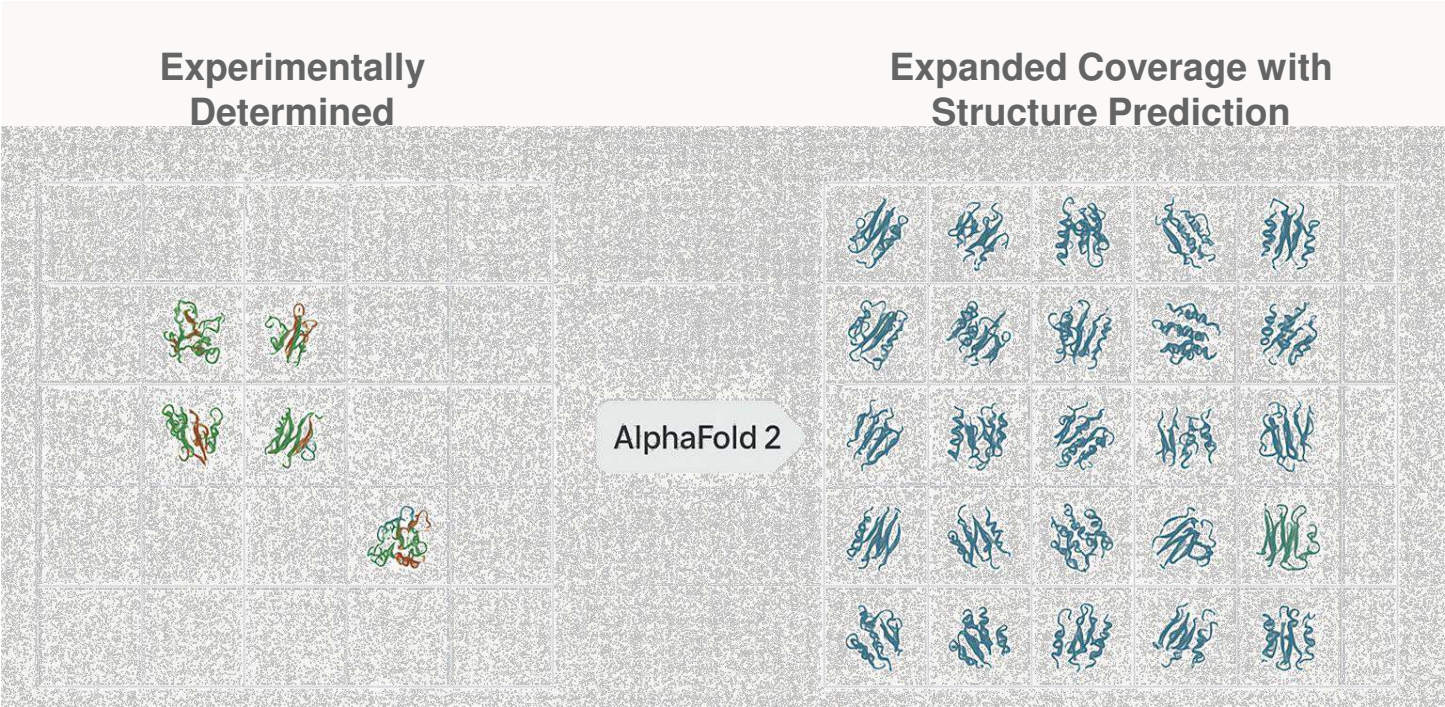


注：模型列表可能不全面。来源：Stanford RAISE Health, 通过 Nestor Maslej et al., 'The AI Index 2025 Annual Report,' AI Index Steering Committee, Stanford HAI (4/25)



Next AI Use Case Frontier – Protein Sequencing = Synthetically Generated Protein Data Yields 1,000x Expansion via AlphaFold

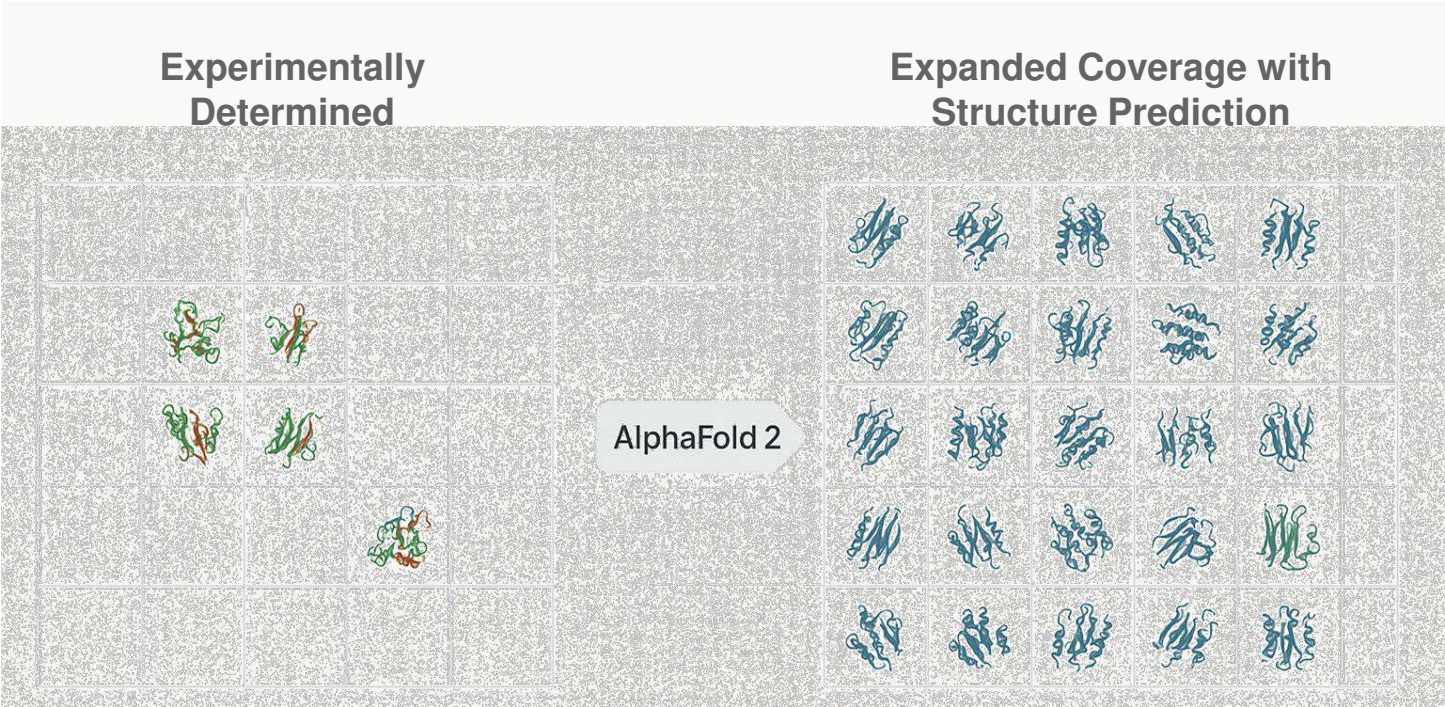
下一个 AI 用例前沿 – 蛋白质测序 =通过 AlphaFold，合成生成的蛋白质数据产量扩大 1,000 倍



214,121  
Protein Structures  
in PDB (2024)



214MM  
Predicted Protein Structures  
in AFDB (2024)



214,121  
Protein Structures  
in PDB (2024)



214MM  
Predicted Protein Structures  
in AFDB (2024)

- 1 Seem Like Change Happening Faster Than Ever?  
Yes, It Is
- 2 AI User + Usage + CapEx Growth =  
Unprecedented
- 3 AI Model Compute Costs High / Rising + Inference Costs Per Token Falling =  
Performance Converging + Developer Usage Rising
- 4 AI Usage + Cost + Loss Growth =  
Unprecedented
- 5 AI Monetization Threats =  
Rising Competition + Open-Source Momentum + China's Rise
- 6 AI & Physical World Ramps =  
Fast + Data-Driven
- 7 Global Internet User Ramps Powered by AI from Get-Go =  
Growth We Have Not Seen Likes of Before
- 8 AI & Work Evolution =  
Real + Rapid

- 1 变化似乎比以往任何时候都快？是的，确实如此
- 2 AI 用户 + 使用量 + 资本支出增长 = 前所未有
- 3 AI 模型计算成本高 / 上升 + 每次 Token 的推断成本下降 = 性能趋同 + 开发者使用量上升
- 4 AI 使用量 + 成本 + 损失增长 = 前所未有
- 5 AI Monetization Threats =  
竞争加剧 + 开放源代码势头 + 中国崛起
- 6 AI 与物理世界加速发展 = 快速 + 数据驱动
- 7 Global Internet User Ramps Powered by AI from Get-Go = 增长是我们前所未见的
- 8 AI 与工作演变 = 真实 + 迅速

*AI Monetization Threats*

=

***Rising Competition***

+

*Open-Source Model Momentum*  
*(& China's Rise)*

AI 货币化威胁 =

**竞争加剧 +**

*Open-Source Model Momentum*  
(与中国的崛起)



Rising Competition =  
AI Model Releases

竞争加剧 =  
AI Model Releases

On the back of Google’s ‘Attention is All You Need’ Transformers research paper in 2017, the first wave of ‘modern AI’ (read: LLMs) focused on text: models such as OpenAI’s GPT-3 and Meta’s Llama-1 showed that teaching computers to finish sentences at scale could unlock broad reasoning abilities. Yet human communication is rarely text-only, and often not even text-first. Images, audio, video, and sensor readings carry context that words alone miss, so researchers at the same companies – and peers like Google, Anthropic, and xAI, among others – began extending language models to handle additional signals.

**Multimodal AI** models are the result. They embed text, pictures, sound, and video into a shared representation and generate outputs in any of those formats. A single query can reference a paragraph and a diagram, and the model can respond with a spoken summary or an annotated image – without switching systems. Each new modality forces models to align meaning across formats rather than optimize for one.

The path to this capability unfolded stepwise: OpenAI’s CLIP paired vision and language in 2021; Meta followed with ImageBind in 2023 and Chameleon in 2024; and by 2024-2025, frontier systems such as GPT-4o, Claude 3, and Chameleon had become fully multimodal. Each new modality forced the models to align meaning across formats rather than optimize for one.

The payoff is practical. A field engineer can aim a phone camera at machinery and receive a plain-language fault diagnosis; a clinician can attach an X-ray to a note and get a structured report draft; and an analyst can combine charts, transcripts, and audio clips in a single query. Compared with text-only models, multimodal systems cut context switching, capture richer detail, and enable applications – quality control, assistive tech, content creation – where visual or auditory information matters as much as words.

On the back of Google’s ‘Attention is All You Need’ Transformers research paper in 2017, the first wave of ‘modern AI’ (read: LLMs) focused on text: models such as OpenAI’s GPT-3 and Meta’s Llama-1 showed that teaching computers to finish sentences at scale could unlock broad reasoning abilities. Yet human communication is rarely text-only, and often not even text-first. I图像、音频、视频和传感器读数带有单词本身所不具备的上下文信息 ss, 因此，在同一公司的研究人员 – 例如 and peeGoogle、Anthropic 和 xAI 等 others – began extending language models to handle additional signals.

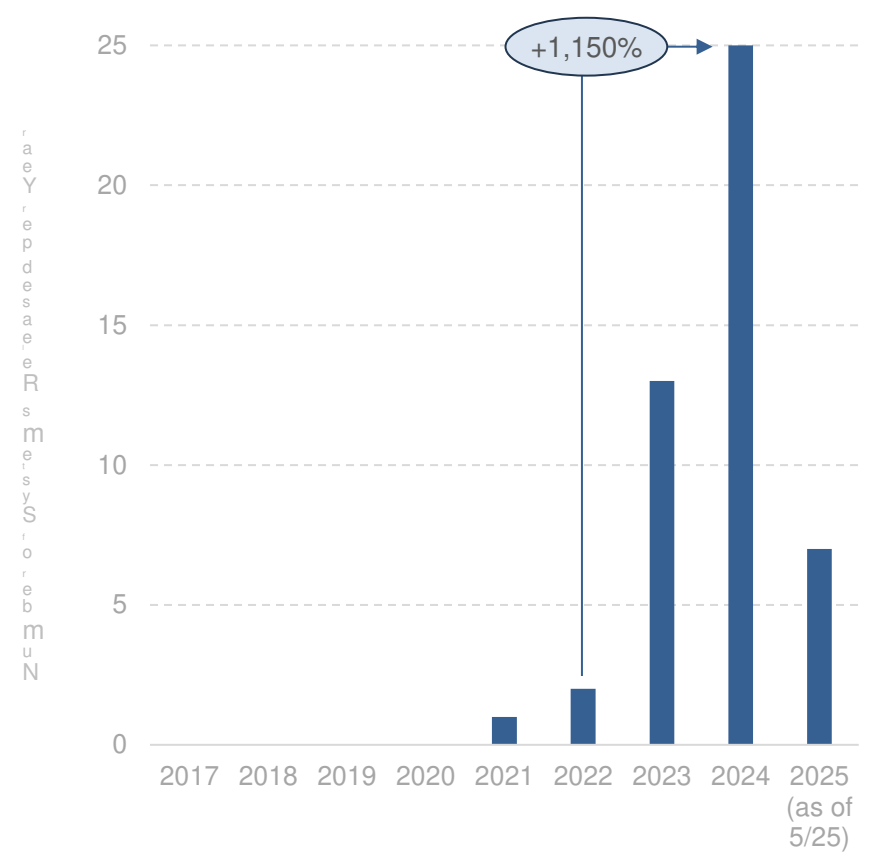
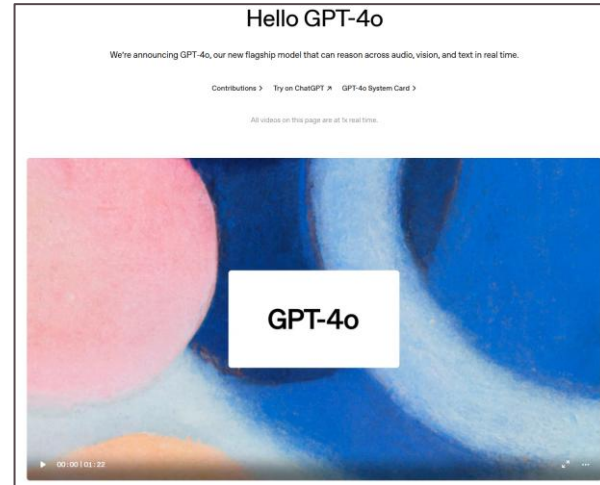
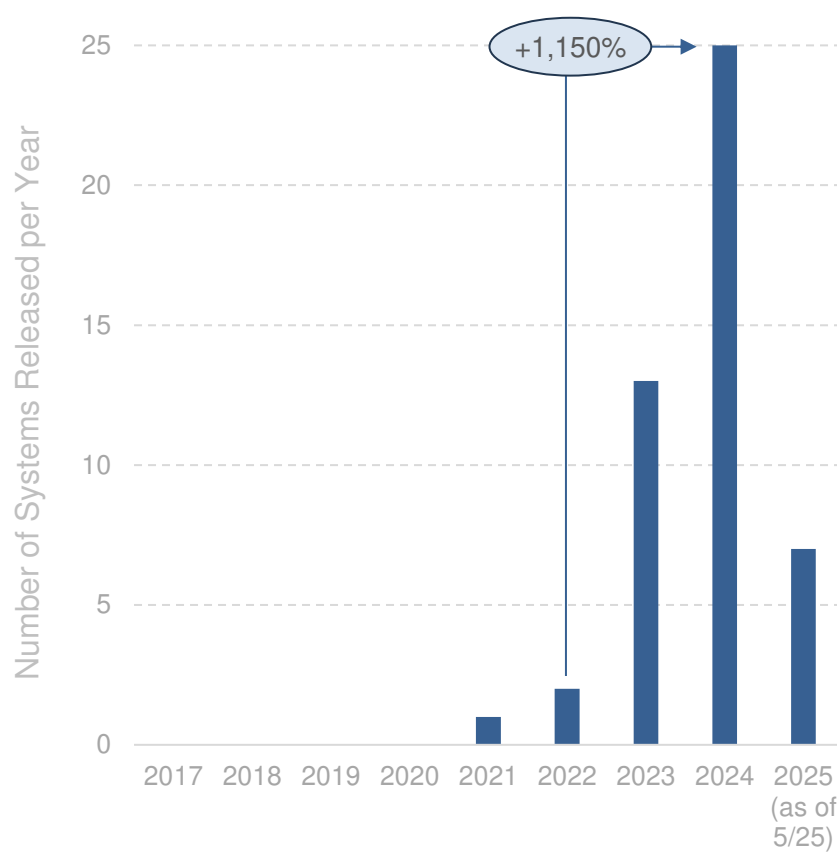
多模态 AI 模型是其结果。它们将文本、图片、声音和视频嵌入到共享表示中，并以任何这些格式生成输出。单个查询可以引用一段文字和一个图表，模型可以用口头摘要或带注释的图像作为响应 – 无需切换系统。每种新的模态都迫使模型对齐跨格式的含义，而不是针对一种格式进行优化。

实现这种能力的路径是逐步展开的：OpenAI 的 CLIP 在 2021 年将视觉和语言配对；Meta 紧随其后，分别在 2023 年和 2024 年推出了 ImageBind 和 Chameleon ；到 2024-2025 年，GPT-4o、Claude 3 和 Chameleon 等前沿系统已完全实现多模态。每种新的模态都迫使模型对齐跨格式的含义，而不是针对一种格式进行优化。

The payoff is practical. 现场工程师可以将手机摄像头对准机械设备，并收到简单的故障诊断；临床医生可以将 X 光片附加到便笺上，并获得结构化的报告草稿；分析师可以在单个查询中组合图表、文字记录和音频剪辑。与纯文本模型相比，多模态系统减少了上下文切换，捕获了更丰富的细节，并支持以下应用 – 质量控制、辅助技术、内容创建 – 在这些应用中，视觉或听觉信息与文字同样重要。

Large-Scale AI *Multimodal*\* Model Competition =  
+1,150% Rise in Models Released Over Two Years, per Epoch AI

## Large-Scale\*\* Multimodal Models – Releases

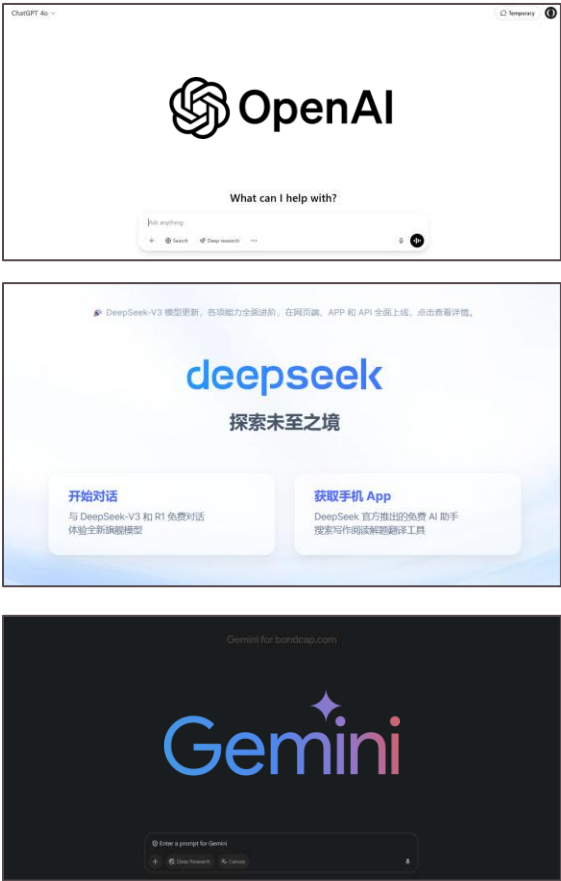


\* 多模态 AI 模型是一种可以处理和整合多种类型数据（例如文本、图像、音频或视频）的模型，以理解和生成跨不同模态的输出。\*\*Epoch AI 将大规模定义为训练计算量经证实超过 10<sup>23</sup> 浮点运算的模型。一个 AI 系统可以在多个领域中运行，并且可能在多个页面中重复计算。来源：Epoch AI 通过 Our World in Data (4/25)、OpenAI、DeepSeek、Google

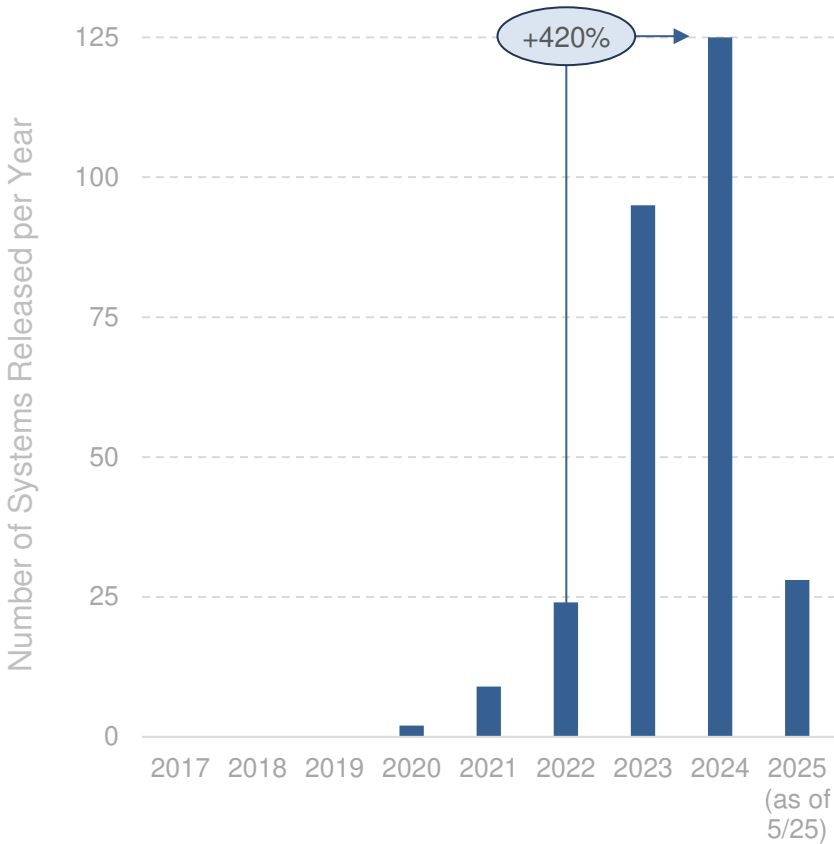
Large-Scale AI *Language* Model Competition = +420% Increase in Models Released Over Two Years, per Epoch AI

Large-Scale AI *Language* Model Competition = +420% Increase in Models Released Over Two Years, per Epoch AI

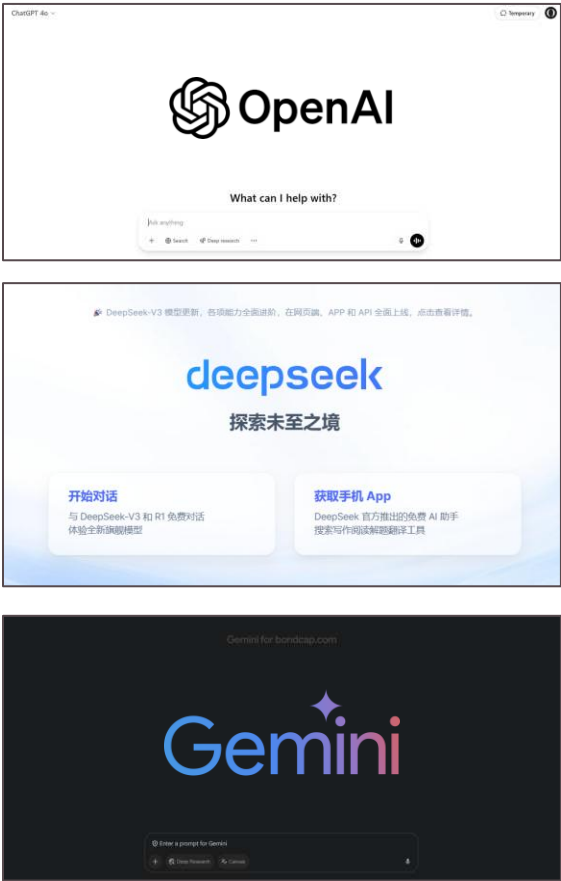
Language Models – Examples



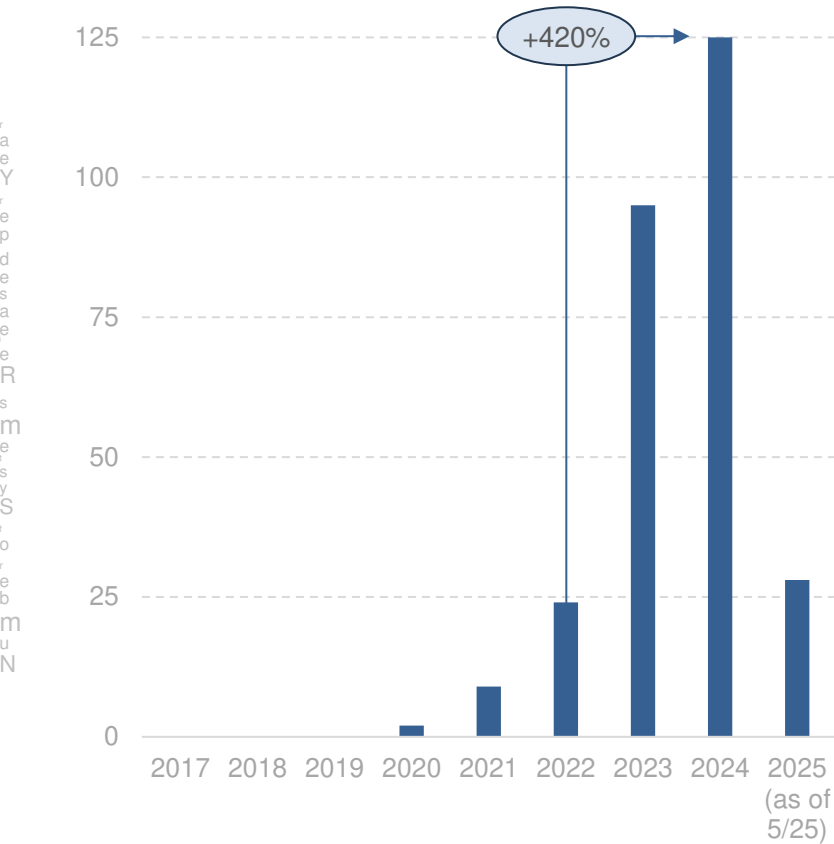
Large-Scale\* Language Models – Releases



Language Models – Examples



大规模 \* 语言模型 – 发布



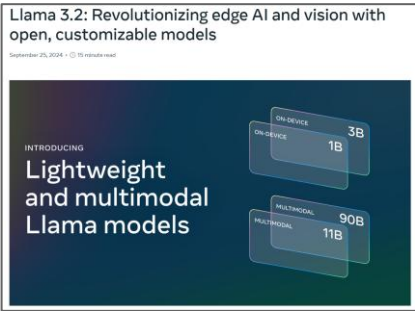
\*Epoch AI defines large-scale as models where their training compute is confirmed to exceed 10<sup>23</sup> floating-point operations. An AI system can operate in more than one domain and may be double-counted across pages. Many models shown are multimodal. Source: Epoch AI via Our World in Data (4/25), OpenAI, DeepSeek, Google

\*Epoch AI 将大规模定义为训练计算量经确认超过 10<sup>23</sup> 次浮点运算的模型。一个 AI 系统可以在多个领域运行，并且可能会在多个页面中重复计算。许多展示模型都是多模态的。来源：Epoch AI 通过 Our World in Data (4/25)、OpenAI、DeepSeek、Google

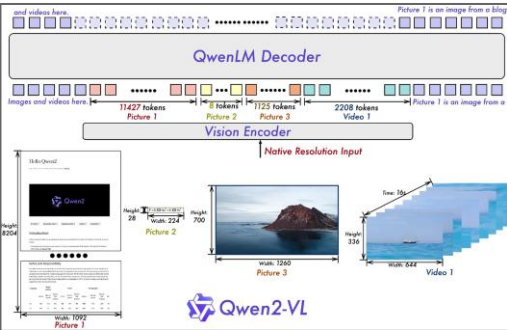
Large-Scale AI *Vision* Model Competition =  
+109% Increase in Models Released Y/Y, per Epoch AI

Large-Scale AI *Vision* Model Competition =  
+109% Increase in Models Released Y/Y, per Epoch AI

Vision Models\* – Examples



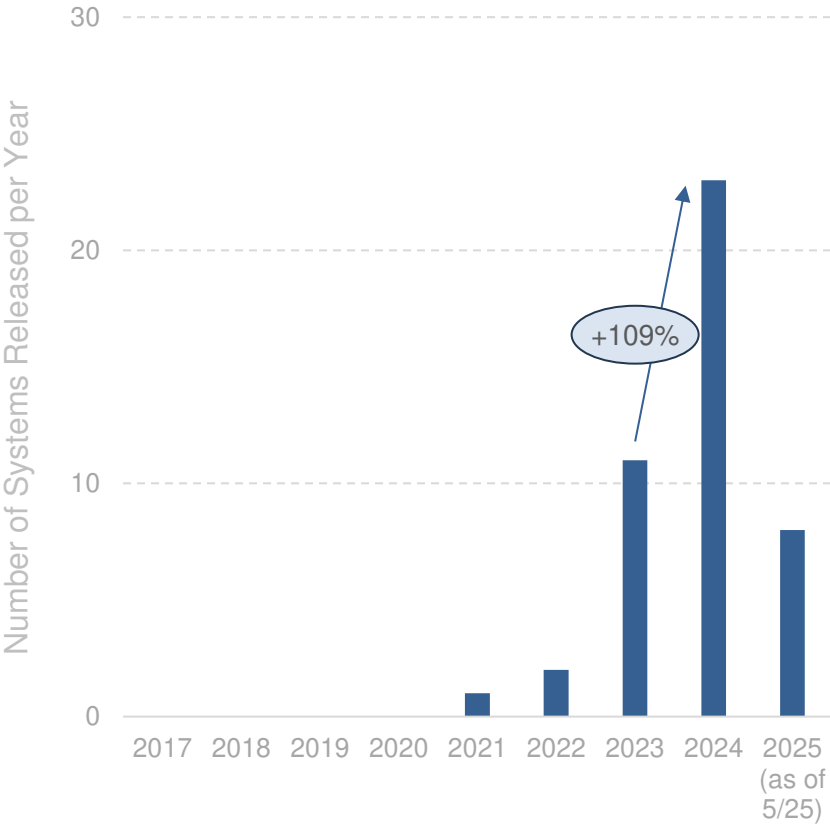
Meta Llama 3.2 – 9/24



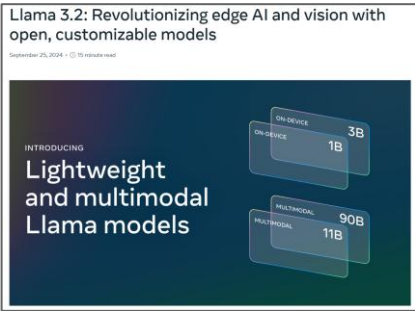
Qwen2-VL – 12/24



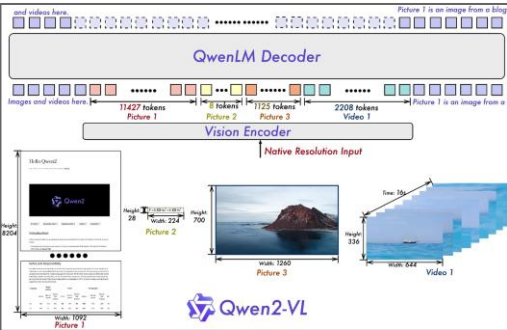
Large-Scale\* Image Models –  
Releases



视觉模型 \* – 示例



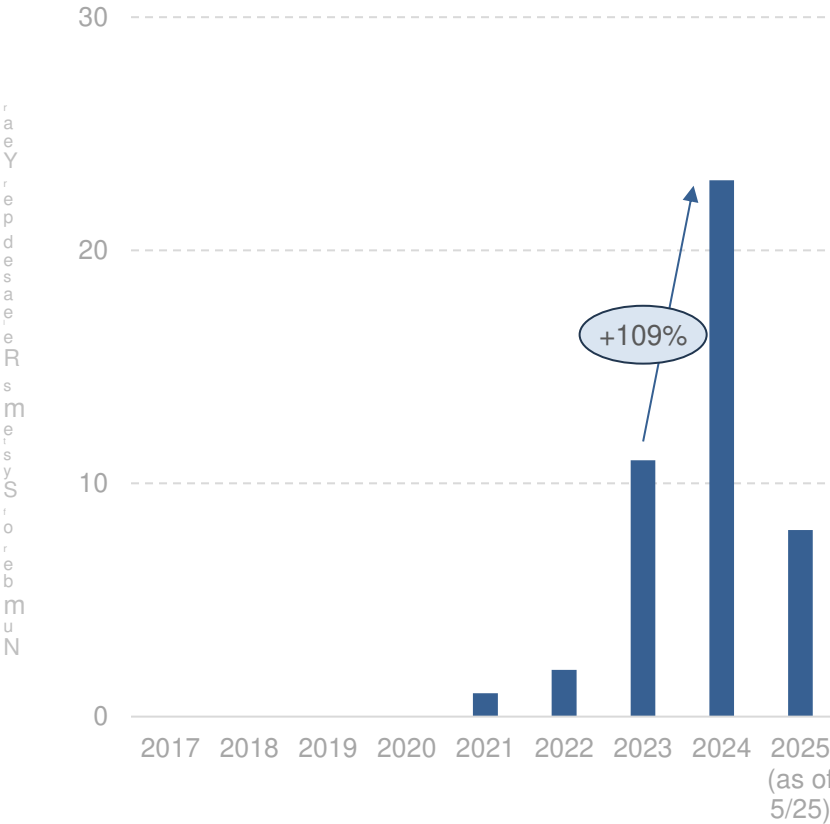
Meta Llama 3.2 – 9/24



Qwen2-VL – 12/24



大规模 \* 图像模型 – 版本



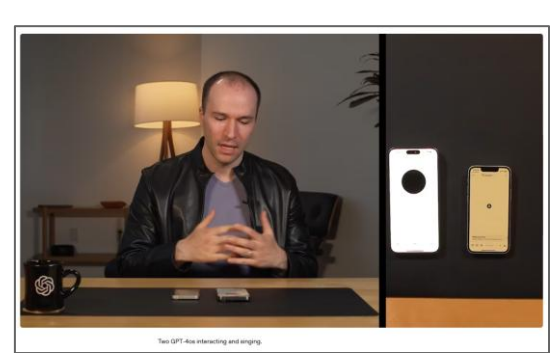
\*Epoch AI defines large-scale as models where their training compute is confirmed to exceed 10<sup>23</sup> floating-point operations. An AI system can operate in more than one domain and may be double-counted across pages. Many models shown are multimodal. Source: Epoch AI via Our World in Data (4/25), Meta, Alibaba

\*Epoch AI 将大规模定义为训练计算量经证实超过 10<sup>23</sup> 次浮点运算的模型。一个 AI 系统可以在多个领域中运行，并且可能会在多个页面中重复计算。显示的许多模型都是多模态的。来源：Epoch AI，通过 Our World in Data (4/25)、Meta、阿里巴巴

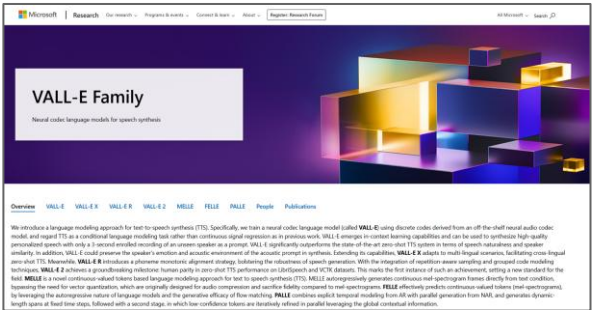
Large-Scale AI *Speech / Audio* Model Competition = +367% Increase in Models Released Y/Y, per Epoch AI

大规模 AI 语音 / 音频模型竞赛 =+367 每年发布的模型数量增加百分比，来源：Epoch AI

Speech / Audio Models – Examples

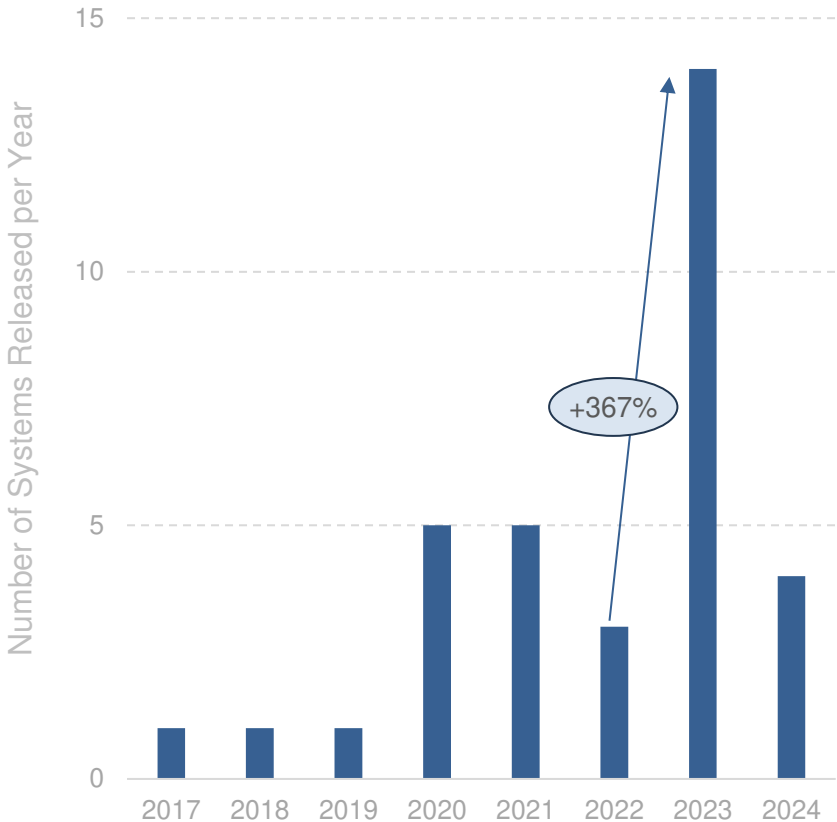


OpenAI GPT 4o Speech – 5/24

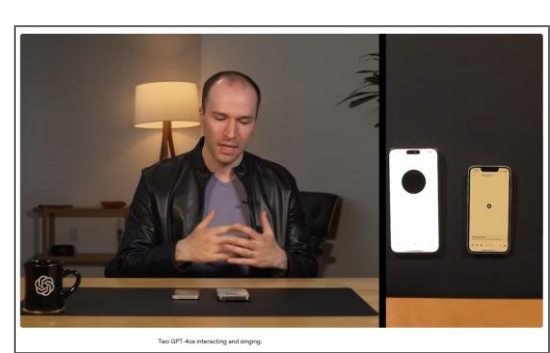


Microsoft VALL-E – 1/23

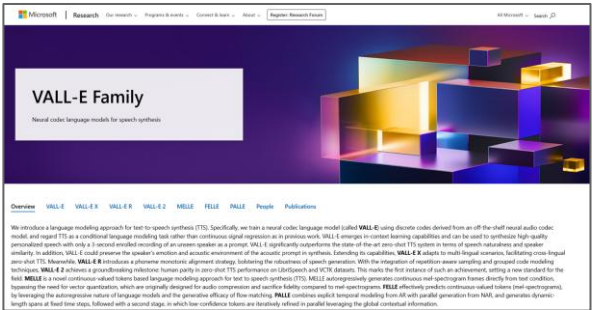
Speech / Audio Models – Releases



语音 / 音频模型 – 示例

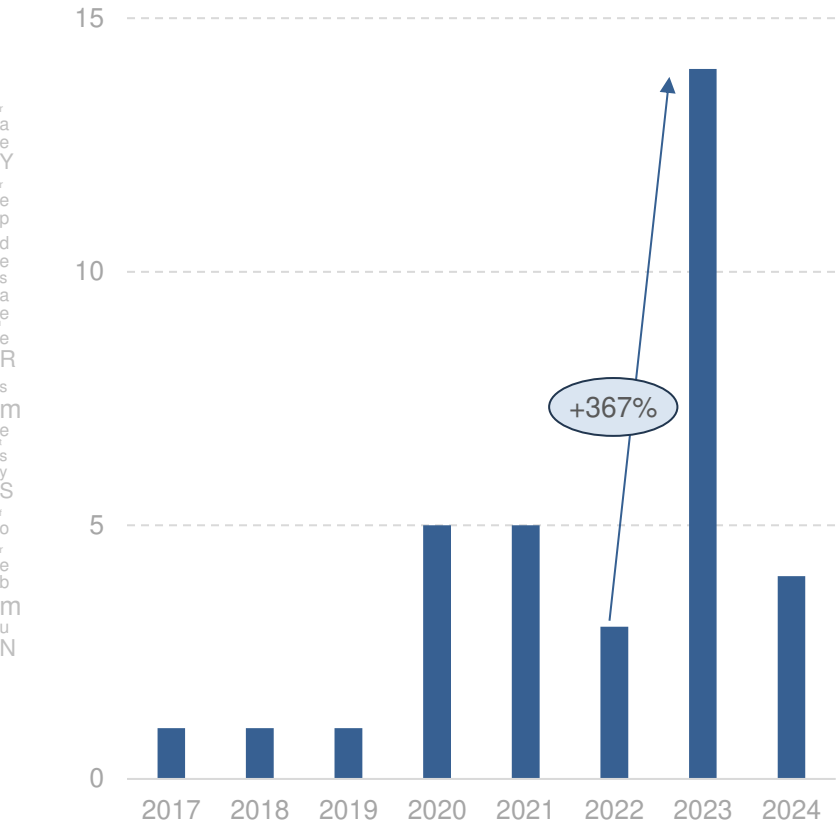


OpenAI GPT 4o 语音 – 5/24



Microsoft VALL-E – 1/23

Speech / Audio Models – Releases



Note: An AI system can operate in more than one domain and may be double-counted across pages. Includes models without verified training compute. Many models shown are multimodal. Source: Epoch AI (5/25), Microsoft (1/23), OpenAI (5/24), Amazon, Pinterest

注意：一个 AI 系统可以在多个领域中运行，并且可能在多个页面中被重复计算。包括未经核实训练计算的模型。显示的许多模型是多模态的。来源：Epoch AI (5/25), Microsoft (1/23), OpenAI (5/24), Amazon, Pinterest

Large-Scale AI Video Model Competition =  
+120% Increase in Models Released Y/Y, per Epoch AI

Large-Scale AI Video Model Competition =  
+120% Increase in Models Released Y/Y, per Epoch AI

Video Models – Examples



OpenAI Sora – 12/24

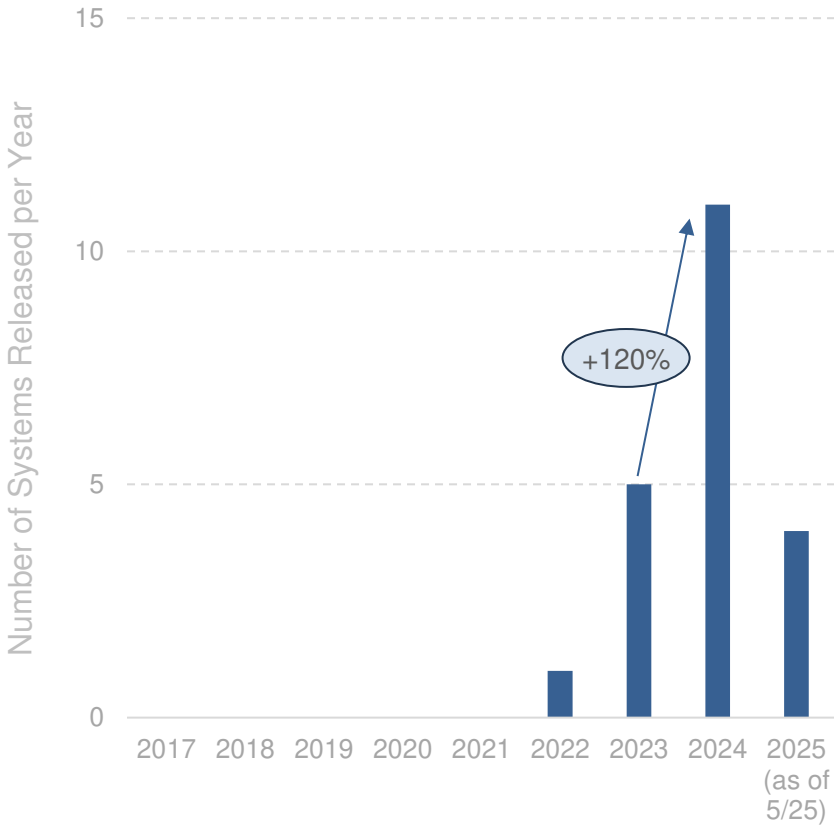


Amazon Nova Reel – 12/24

According to academic studies, 50% of the human brain is wired for visual processing. The ability for users to explore their interest visually and take action on them... is particularly relevant for Gen Z... who have been raised on an internet of visual content across images and video.

- Pinterest CEO Bill Ready (5/25)

Large-Scale\* Video Models – Releases



Video Models – Examples



OpenAI Sora – 12/24

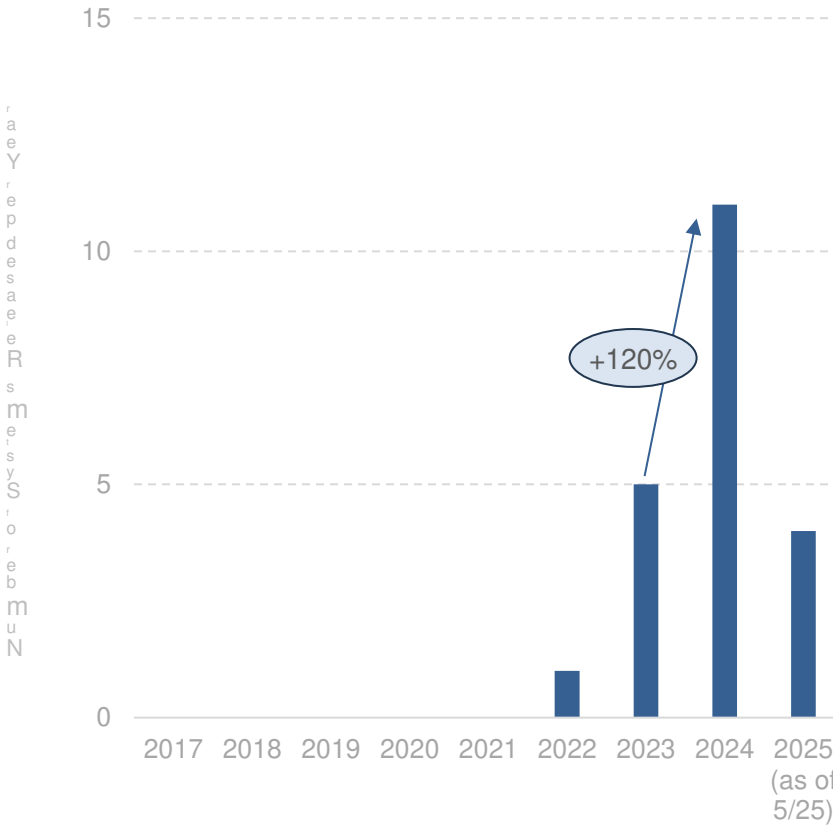


Amazon Nova Reel – 12/24

根据学术研究，人脑有 50% 的功能用于视觉处理。用户通过视觉方式探索他们的兴趣并对其采取行动的能力 ..... 对于在由图像和视频组成的视觉互联网中成长起来的 Z 世代尤其重要。

- Pinterest CEO Bill Ready (5/25)

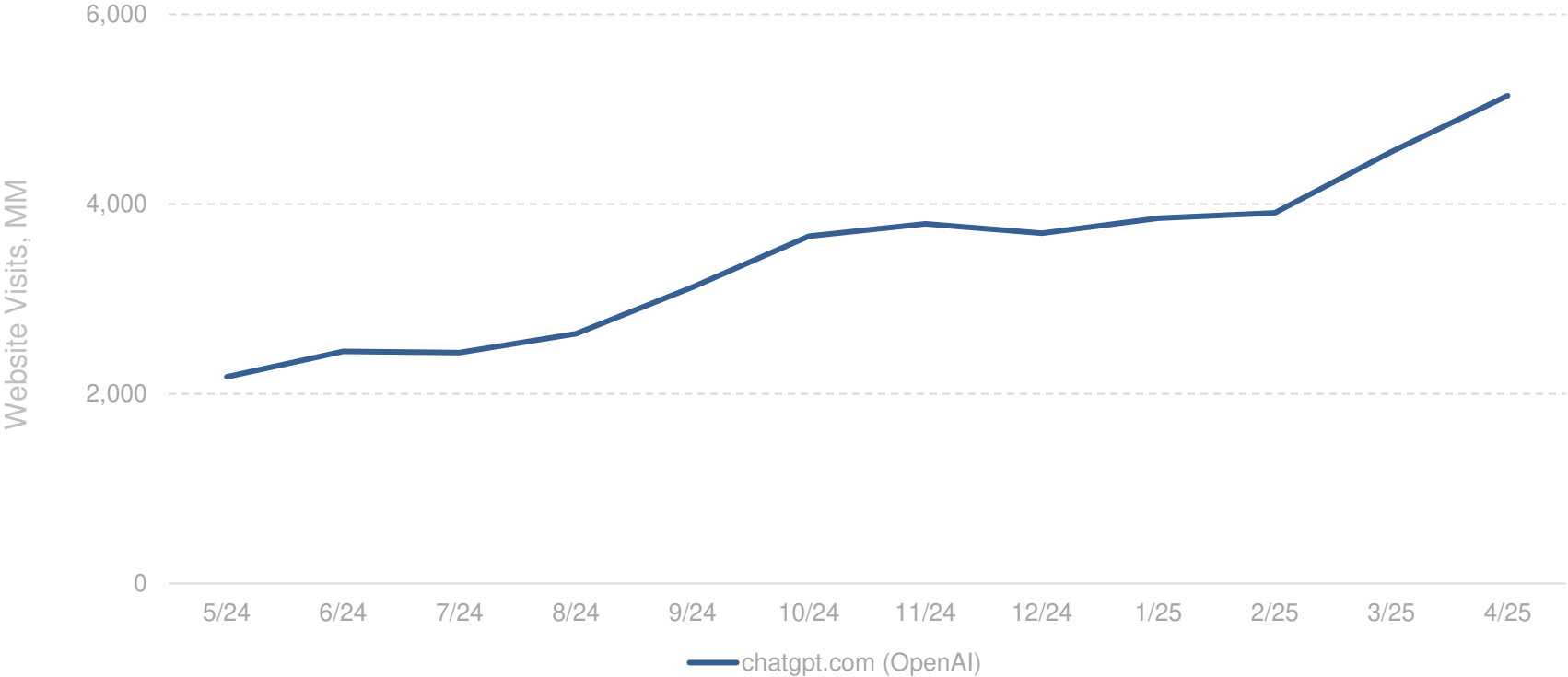
大规模 \* 视频模型 – 发布



\*Epoch AI 将大规模定义为训练计算量经确认超过 10<sup>23</sup> 次浮点运算的模型。一个 AI 系统可以在多个领域运行，并且可能在多个页面中被重复计算。许多展示的模型是多模态的。来源：Epoch AI 通过 Our World in Data (4/25)、OpenAI、Amazon、Pinterest、Pinterest

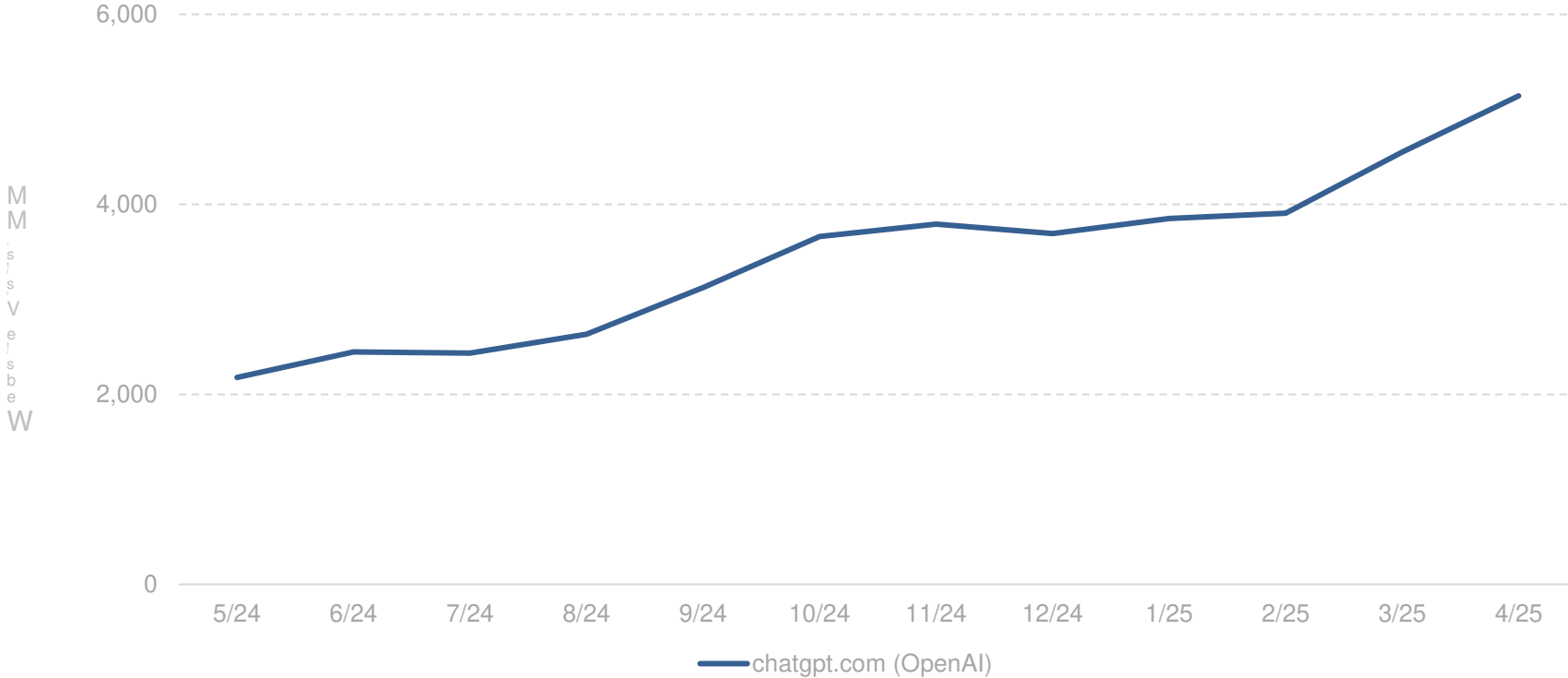


OpenAI ChatGPT Global Website Visits (MM) – 5/24-4/25, per Similarweb



Note: Includes desktop & mobile (non-app) website visits. China data may be subject to informational limitations due to government restrictions. Source: Similarweb (5/25)

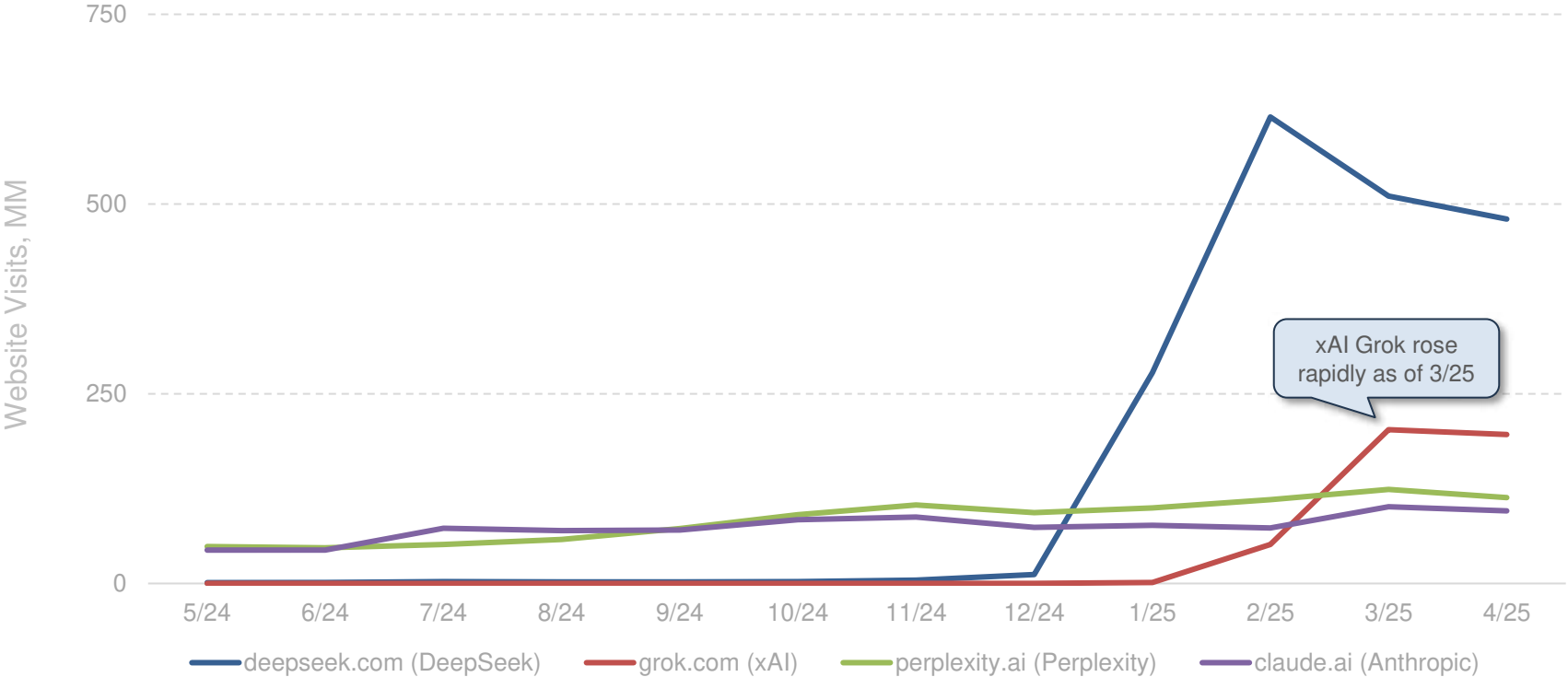
OpenAI ChatGPT Global Website Visits (MM) – 5/24-4/25, per Similarweb



Note: Includes desktop & mobile (non-app) website visits. China data may be subject to informational limitations due to government restrictions. Source: Similarweb (5/25)

...LLM Competition – Website Visits =  
DeepSeek & xAI Grok Also Rising @ 196-480MM Visits Each

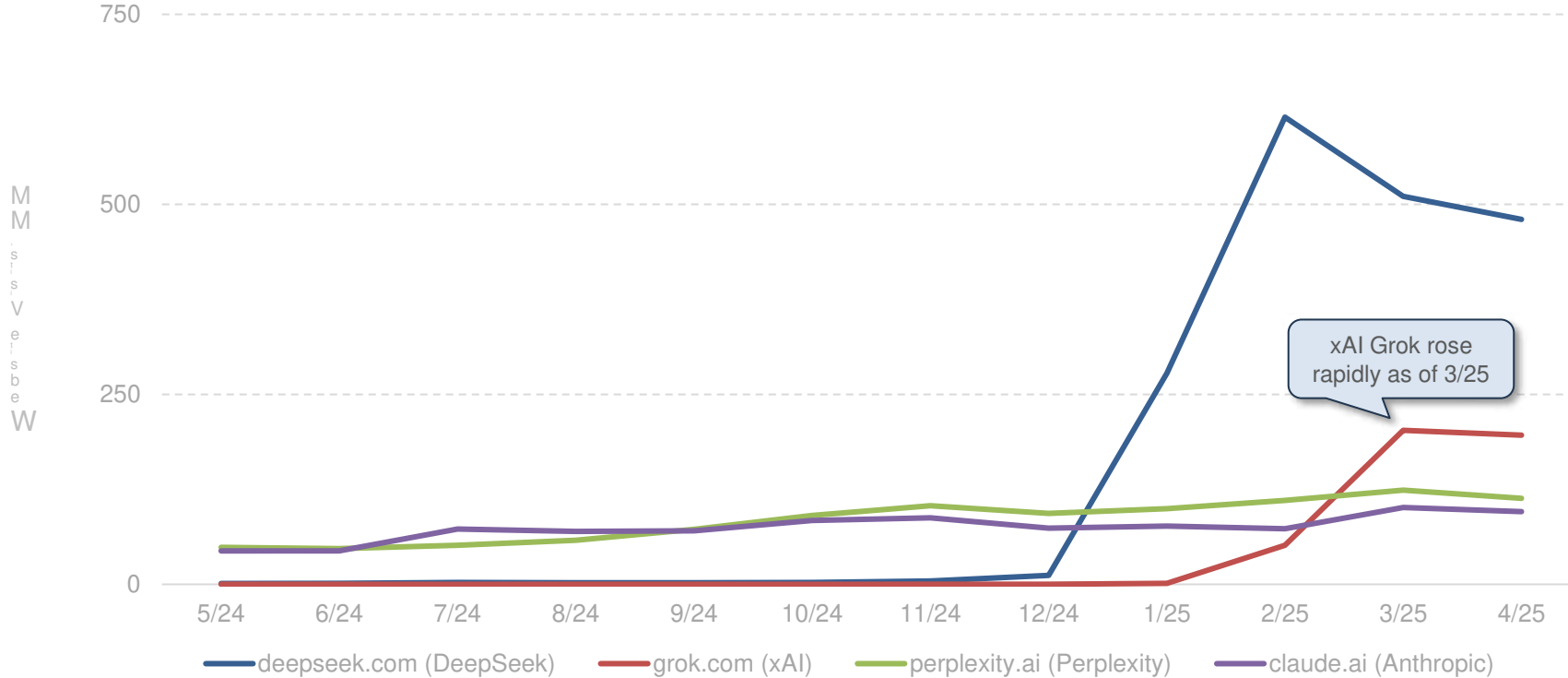
DeepSeek, xAI Grok, Perplexity & Anthropic Claude Global Website Visits (MM) –  
5/24-4/25, per Similarweb



Note: Includes desktop & mobile (non-app) website visits. China data may be subject to informational limitations due to government restrictions. Source: Similarweb (5/25)

...LLM Competition – 网站访问量 =DeepSeek 和 xAI Grok 的访问量也在上升，每个网站的访问量为 1.96-4.8 亿

DeepSeek、xAI Grok、Perplexity 和 Anthropic Claude 全球网站访问量（百万） –  
5/24-4/25, 数据来源: Similarweb



Note: Includes desktop & mobile (non-app) website 访问量。由于政府限制，中国数据可能受到信息限制。来源: Similarweb ( 5/25 )

Select AI Product Announcements – 5/19/25-5/23/25,  
per Google, Microsoft, Anthropic & OpenAI



- Gemini Live camera & screen sharing
- Project Mariner computer use
- Updated Gemini 2.5 Flash
- Gemini 2.5 Pro
- Native audio output for 2.5 Flash & Pro Previews
- Thinking Budgets for Gemini 2.5 Pro
- Deep Think
- Project Astra capabilities
- Gemini in Chrome
- Deep Research improvements
- Gemini Agent Mode
- Google AI Pro Subscription
- Google AI Ultra Subscription
- Google Beam
- Google Meet speech translation
- Personalized Smart Replies
- Jules
- Imagen 4
- Veo 3
- Lyria 2
- Flow TV
- Project Moohan
- Glasses with Android XR
- ‘Try on’ experiment
- Agentic checkout
- Gemini interactive quizzes
- Canvas Create menu
- LearnLM integration into Gemini 2.5
- SDK support for Model Context Protocol (MCP) definitions in Gemini API
- Gemini Diffusion
- SynthID Detector
- Conversational tutor prototype
- Google Live API audiovisual input & native audio out dialogue
- Gemma 3n
- AI studio enhancements
- Android Studio Journeys
- Android Studio Version Upgrade Agent
- Wear OS 6 Developer Preview
- Gemini Code Assist
- New Firebase features
- Google AI Edge Portal
- Google Vids
- Enhanced Audio Overviews
- Sparkify experiment



- Magentic-UI
- Copilot Studio multi-agent orchestration
- GitHub Copilot asynchronous functioning
- Azure AI Foundry expansion
- NLWeb
- Model Context Protocol (MCP) integration
- Entra Agent ID
- SQL Server 2025
- Windows Subsystem for Linux Open-Source
- GitHub Copilot Chat Extension
- Aurora AI-Powered Weather Forecasting



- Claude Opus 4
- Claude Sonnet 4



- Acquisition of io

精选 AI 产品公告 – 5/19/25-5/23/25， 来自 Google、  
Microsoft、Anthropic 和 OpenAI



- Gemini Live 摄像头和屏幕共享
- Project Mariner 计算机使用
- 更新版 Gemini 2.5 Flash
- Gemini 2.5 Pro
- 2.5 Flash 和 Pro 预览版的原生音频输出
- Gemini 2.5 Pro 的思维预算
- 深度思考
- Project Astra 功能
- Chrome 中的 Gemini
- 深度研究改进
- Gemini 代理模式
- Google AI Pro 订阅
- Google AI Ultra 订阅
- Google Beam
- Google Meet 语音翻译
- 个性化智能回复
- 朱尔斯
- Imagen 4
- Veo 3
- Lyria 2
- Flow TV
- Project Moohan
- 带有 Android XR 的眼镜
- ‘Try on’ experiment
- Agentic checkout
- Gemini interactive quizzes
- Canvas Create menu
- LearnLM integration into Gemini 2.5
- SDK support for Model Context Protocol (MCP) definitions in Gemini API
- Gemini Diffusion
- SynthID Detector
- Conversational tutor prototype
- Google Live API audiovisual input & native audio out dialogue
- Gemma 3n
- AI studio enhancements
- Android Studio Journeys
- Android Studio Version Upgrade Agent
- Wear OS 6 Developer Preview
- Gemini Code Assist
- New Firebase features
- Google AI Edge Portal
- Google Vids
- Enhanced Audio Overviews
- Sparkify experiment



- Magentic-UI
- Copilot Studio 多代理业务流程
- GitHub Copilot 异步功能
- Azure AI Foundry 扩展
- NLWeb
- 模型上下文协议 (MCP) 集成
- Entra 代理 ID
- SQL Server 2025
- 适用于 Linux 的 Windows 子系统开源
- GitHub Copilot Chat Extension
- Aurora AI 驱动的天气预报



- Claude Opus 4
- Claude Sonnet 4



- 收购 io

Note: Announcements include products that were made immediately-available and forthcoming products. List is non-exhaustive. Source: Google Microsoft, Anthropic, OpenAI (5/25)

Note: Announcements include products that were made immediately-available and forthcoming products. List is non-exhaustive. Source: Google Microsoft, Anthropic, OpenAI (5/25)

*AI Monetization Threats*  
=  
*Rising Competition*  
+  
*Open-Source Model Momentum*  
*(& China's Rise)*

AI 货币化威胁 (及中国崛起)  
=+  
\_\_\_\_\_

To understand where AI model development is headed, it helps to examine how two distinct approaches – closed-source and open-source – have evolved and diverged.

In the early days of modern machine learning (2012-2018), most models were open-source, rooted in academic and collaborative traditions.

But as AI systems became more powerful and commercially valuable, and as development shifted from academia to industry, a parallel movement emerged – around 2019 (when GPT-2 launched with restricted weights), the development of proprietary (closed-source) models, motivated by proprietary interests, competitive advantage, and safety concerns.

Closed models follow a centralized, capital-intensive arc. These models – like OpenAI’s GPT-4 or Anthropic’s Claude – are trained within proprietary systems on massive proprietary datasets, requiring months of compute time and millions in spending. They often deliver more capable performance and easier usability, and thus are preferred by enterprises and consumers, and – increasingly – governments. However, the tradeoff is opacity: no access to weights, training data, or fine-tuning methods.

What began as a research frontier became a gated product experience, served via APIs, licensed to enterprises, and defended by legal and commercial firewalls. Now, the AI race is coming full circle.

As LLMs mature – and competition intensifies – we are seeing resurgence of open-source models owing to their lower costs, growing capabilities, and broader accessibility for developers and enterprises alike.

These are freely available for anyone to use, modify, and build upon, and thus are generally preferred by early-stage startups, researchers / academics, and independent developers.

Platforms like Hugging Face have made it frictionless to download models like Meta’s Llama or Mistral’s Mixtral, giving startups, academics, and governments access to frontier-level AI without billion-dollar budgets.

Open-source AI has become the garage lab of the modern tech era: fast, messy, global, and fiercely collaborative. And China (as of Q2:25) – based on the number of large-scale AI models\* released – is leading the open-source race, with three large-scale models released in 2025 – DeepSeek-R1, Alibaba Qwen-32B and Baidu Ernie 4.5\*\*.

The split has consequences. Open-source is fueling sovereign AI initiatives, local language models, and community-led innovation. Closed models, meanwhile, are dominating consumer market share and large enterprise adoption.

We’re watching two philosophies unfold in parallel – freedom vs. control, speed vs. safety, openness vs. optimization – each shaping not just how AI works, but who gets to wield it.

*\*Large-scale AI models = Models with training compute confirmed to exceed 10<sup>23</sup> floating point operations.*  
*\*\*To be made open-source as of 6/30/25, per Baidu.*

To understand where AI model development is headed, it helps to examine how two distinct methods – closed and open – are evolving and diverging.

In the early days of modern machine learning (2012-2018), most models were open-source, rooted in academic and collaborative traditions.

But as AI systems became more powerful and commercially valuable, and as development shifted from academia to industry, a parallel movement emerged – around 2019 (when GPT-2 launched with restricted weights), the development of proprietary (closed-source) models, motivated by proprietary interests, competitive advantage, and safety concerns.

封闭模型遵循中心化的、资本密集型的轨迹。这些模型 – 比如 OpenAI 的 GPT-4 或 Anthropic 的 Claude – 在专有系统上，利用大量的专有数据集进行训练，需要数月的计算时间和数百万美元的支出。它们通常提供更强大的性能和更简单的可用性，因此受到企业、消费者以及 – 越来越多的 – 政府的青睐。然而，其代价是不透明：无法访问权重、训练数据或微调方法。

What began as a research frontier became a gated product experience, served via APIs, licensed to enterprises, and defended by legal and commercial firewalls. Now, the AI race is coming full circle.

As LLMs mature – and competition intensifies – we are seeing resurgence of open-source models owing to their lower costs, growing capabilities, and broader accessibility for developers and enterprises alike.

These are freely available for anyone to use, modify, and build upon, and thus are generally preferred by early-stage startups, researchers / academics, and independent developers.

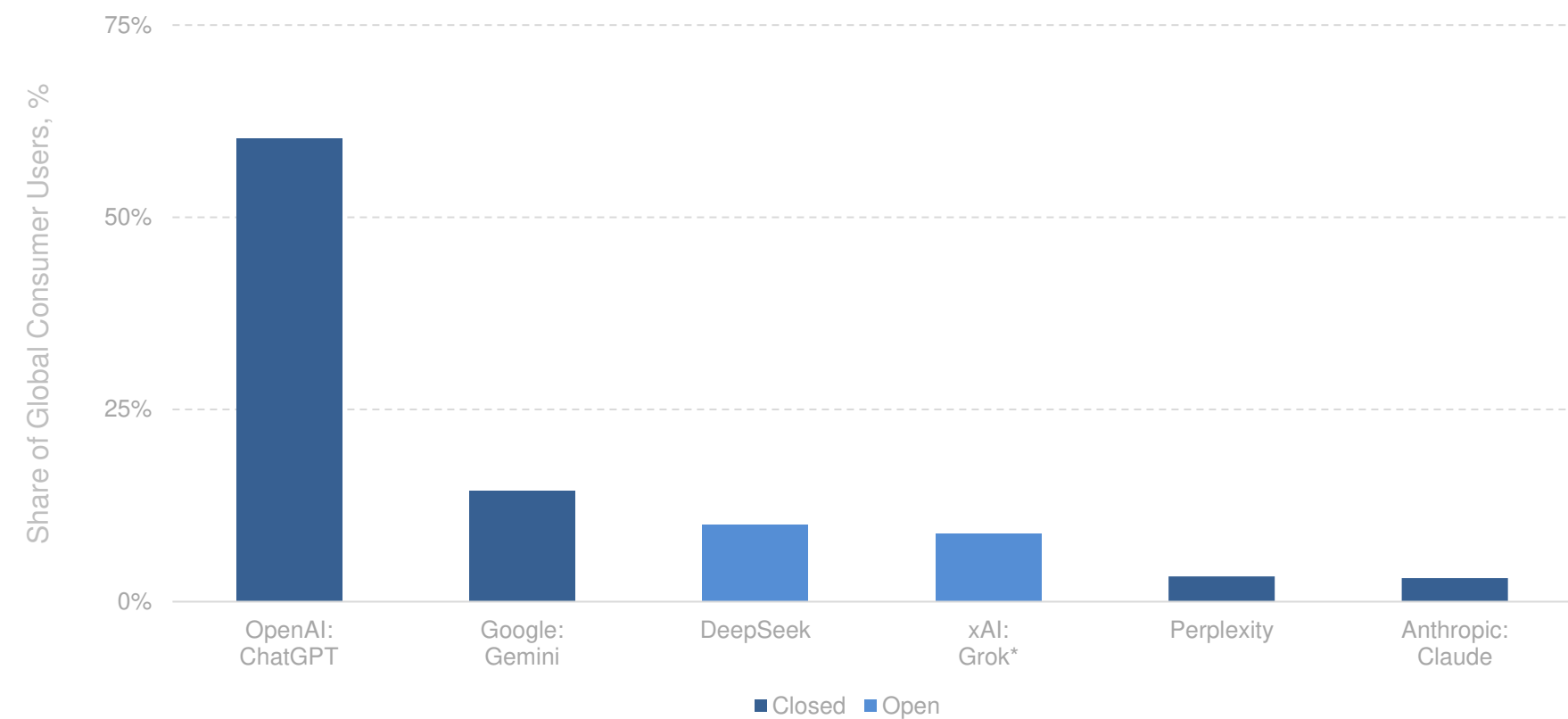
Hugging Face 等平台让下载 Meta 的 Llama 或 Mistral 的 Mixtral 等模型变得非常容易，使初创公司、学者和政府无需数十亿美元的预算即可获得前沿水平的 AI。开源 AI 已成为现代科技时代的车库实验室：快速、混乱、全球化且高度协作。而中国（截至 25 年第二季度） – 基于发布的大型 AI 模型 \* 的数量 – 在开源竞赛中处于领先地位，2025 年发布了三个大型模型 – DeepSeek-R1、阿里巴巴 Qwen-32B 和百度 Ernie 4.5\*\*。

这种分裂会产生后果。开源正在推动主权 AI 计划、本地语言模型和社区主导的创新。与此同时，封闭模型正在主导消费者市场份额和大型企业采用。我们正在关注两种哲学并行发展 —— 自由与控制、速度与安全、开放与优化 —— 每一种哲学不仅塑造了 AI 的工作方式，还塑造了谁能使用它。

*\* 大规模 AI 模型 = 训练计算量经证实超过 10<sup>23</sup> 次浮点运算的模型。 \*\* 根据百度的数据，截至 25 年 6 月 30 日，将开源。*

# Closed vs. Open-Source Models – Monthly Active Users (MAUs) = Closed Models Dominating With Consumers, per YipitData

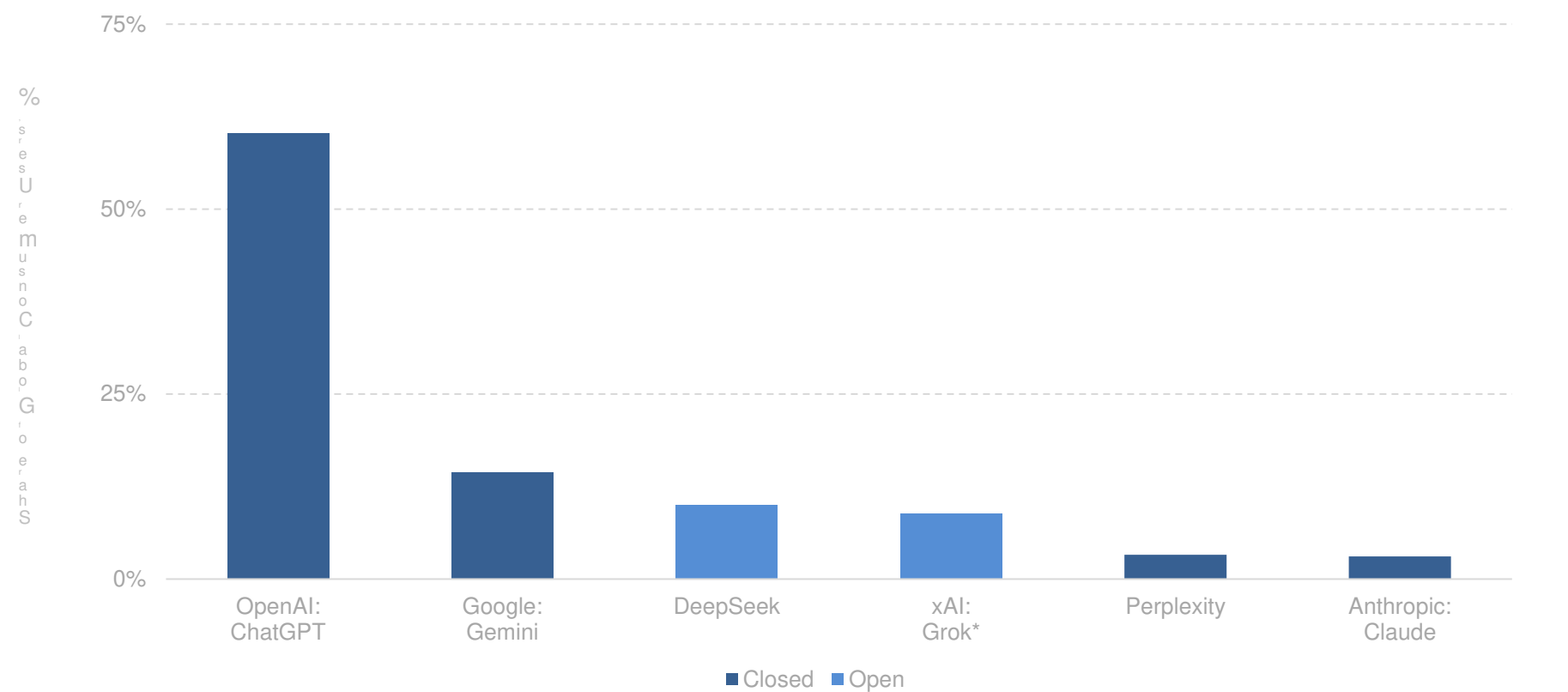
Estimated Share of Global Monthly Active Users (MAUs) Across Six Leading LLMs – 4/25, per YipitData



*\*xAI open-sourced the Grok-1 base model in March 2024, but newer versions and full chatbot features remain proprietary. Note: Data is a subset of global internet users and absolute user data will be understated; however, given that the panel is globally-representative (with limitations on China-specific data), relative comparisons / trends are informative. Desktop users only. Figures calculate the number of users on a given platform, divided by the number of users on all platforms combined. Figures are non-deduped (i.e., users using multiple platforms may be counted twice). Data measures several million global active desktop users' clickstream data. Data consists of users' web requests & is collected from web services / applications, such as VPNs and browser extensions. Panel is globally-representative (with limitations on China-specific data). Users must have been part of the panel for 2 consecutive months to be included. Source: YipitData (5/25)*

# 封闭 vs. 开源模型 – 月活跃用户 (MAU) =YipitData 数据显示封闭模型在消费者中占据主导地位

全球六大领先 LLM 的月活跃用户 (MAU) 估计份额 – YipitData, 4/25

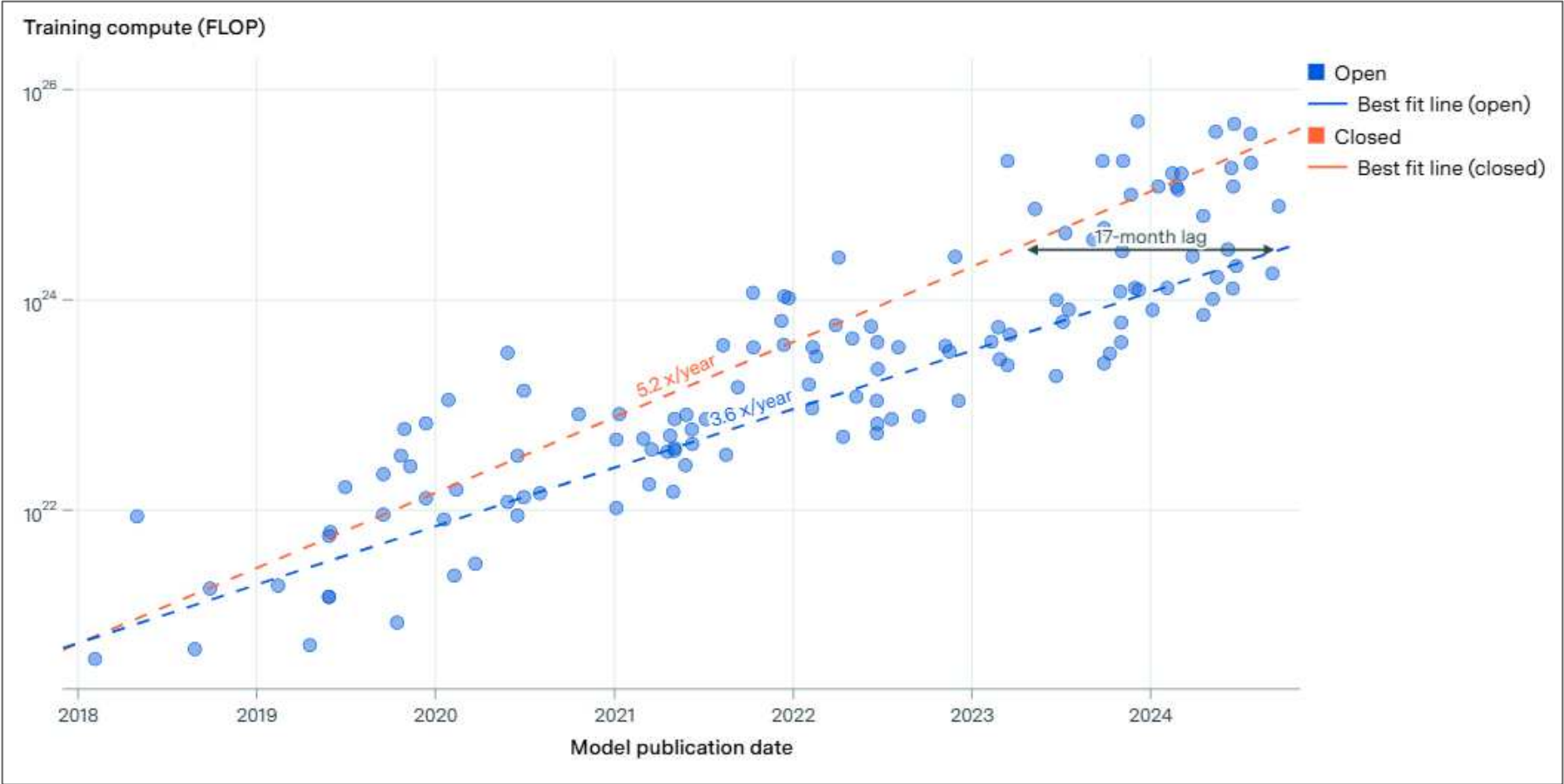


*\*xAI 于 2024 年 3 月开源了 Grok-1 基础模型，但更新的版本和完整的聊天机器人功能仍为专有。注意：数据是全球互联网用户的一个子集，绝对用户数据将被低估；然而，鉴于该小组具有全球代表性（对中国特定数据存在限制），相对比较 / 趋势具有参考意义。仅限桌面用户。数据计算的是给定平台上的用户数量，除以所有平台上用户的总数。数字未去重（即，使用多个平台的用户可能会被计算两次）。数据测量的是数百万全球活跃桌面用户的点击流数据。数据包括用户的网络请求，这些数据是从网络服务 / 应用程序（如 VPN 和浏览器扩展）收集的。该小组具有全球代表性（对中国特定数据存在限制）。用户必须连续 2 个月成为该小组的成员才能被包括在内。来源：YipitData （ 5/25 ）*

# Closed vs. Open-Source Models – Compute Investment = Closed Models Higher, per Epoch AI

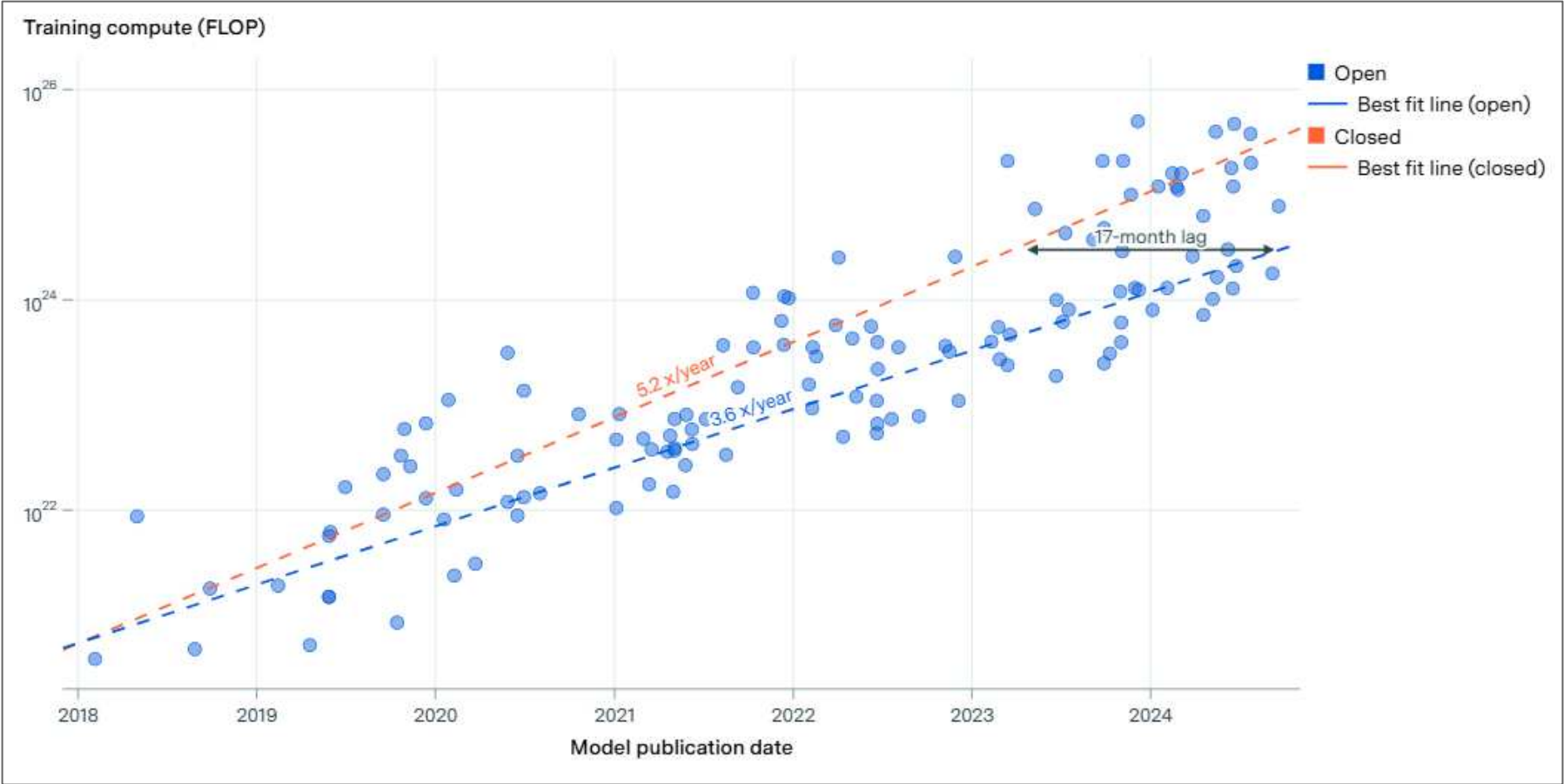
封闭与开源模型 – 计算投资 = 每个 Epoch AI，封闭模型更高

Training Compute Resources for Open vs. Closed LLMs – 2/18-9/24, per Epoch AI



Source: Epoch AI (11/24)

Training Compute Resources for Open vs. Closed LLMs – 2/18-9/24, per Epoch AI

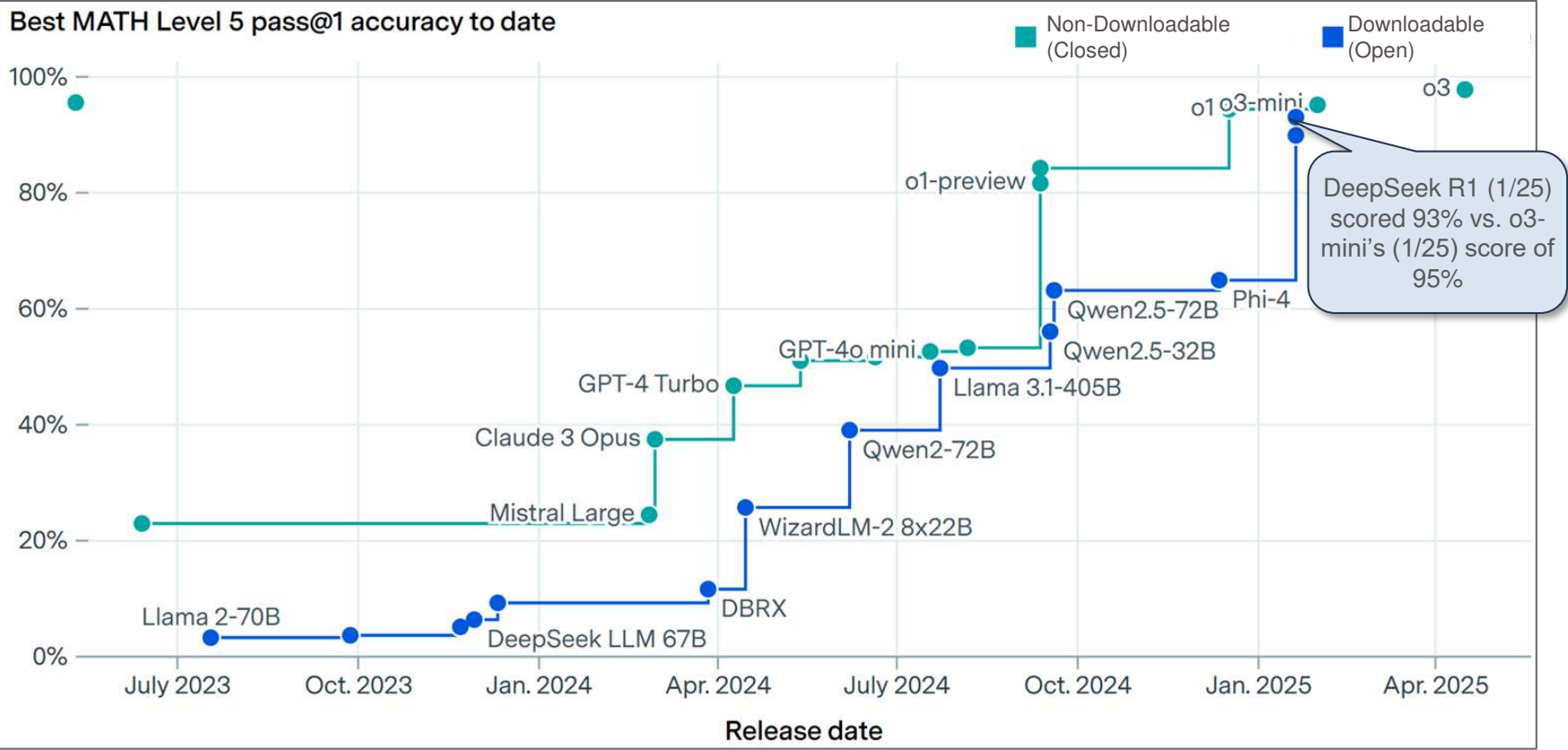


Source: Epoch AI (11/24)



# Closed vs. Open-Source Models – Performance = Gap Closing...China Rising, per Epoch AI...

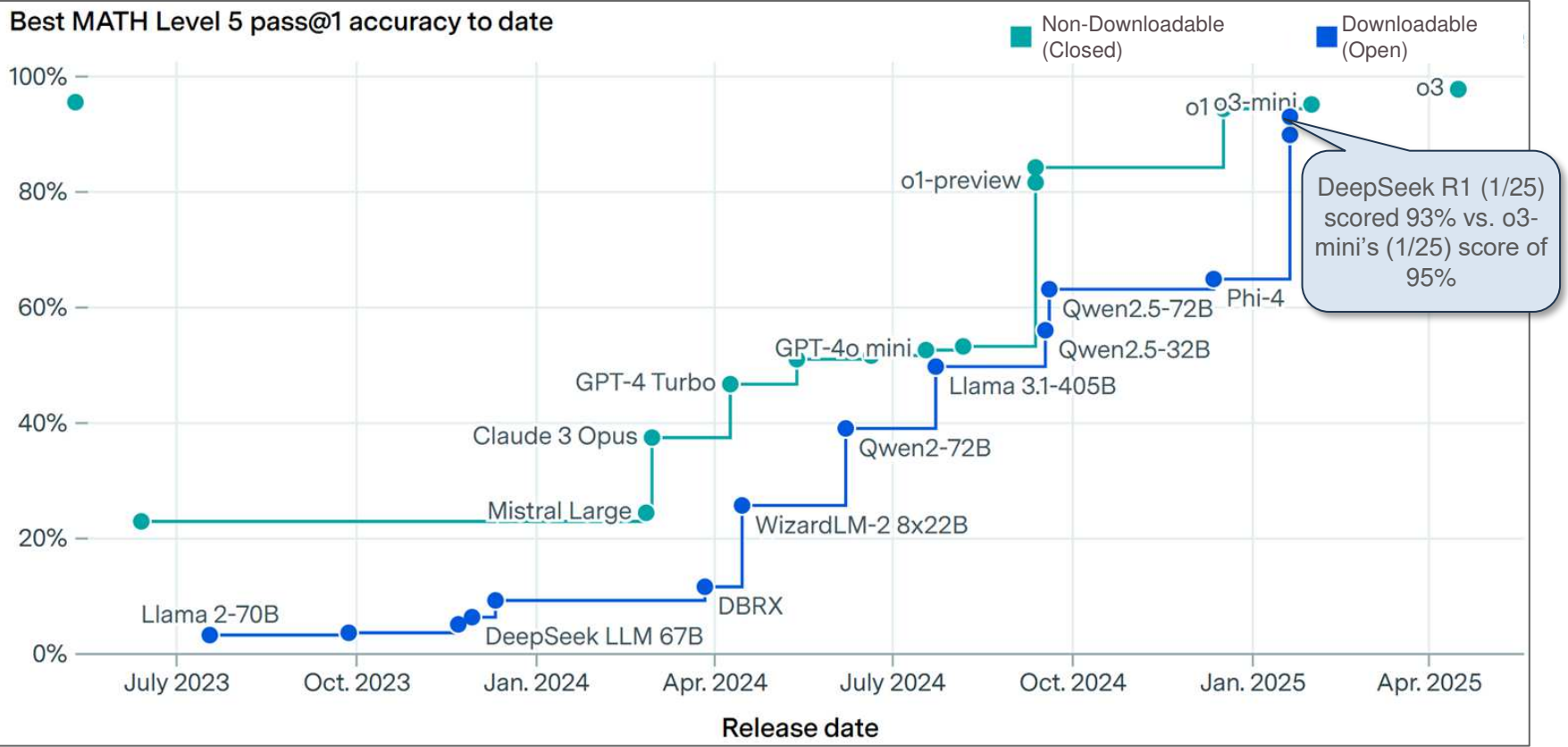
Performance on MATH Level 5 Test, Open vs. Closed LLMs by Year Released – 6/23-4/25, per Epoch AI



Note: MATH Level 5 pass@1 refers to the accuracy of an AI model on the MATH benchmark, a dataset of high school competition-level mathematics problems. Level 5 indicates the most challenging problems in the benchmark. 'pass@1' measures whether the model correctly solves the problem on its first attempt. Source: Epoch AI (5/25)

# 封闭与开源模型 – 性能 = 差距缩小… 中国崛起，根据 Epoch AI…

MATH 5 级测试的性能，按发布年份划分的开放与封闭 LLM – 6/23-4/25，根据 Epoch AI

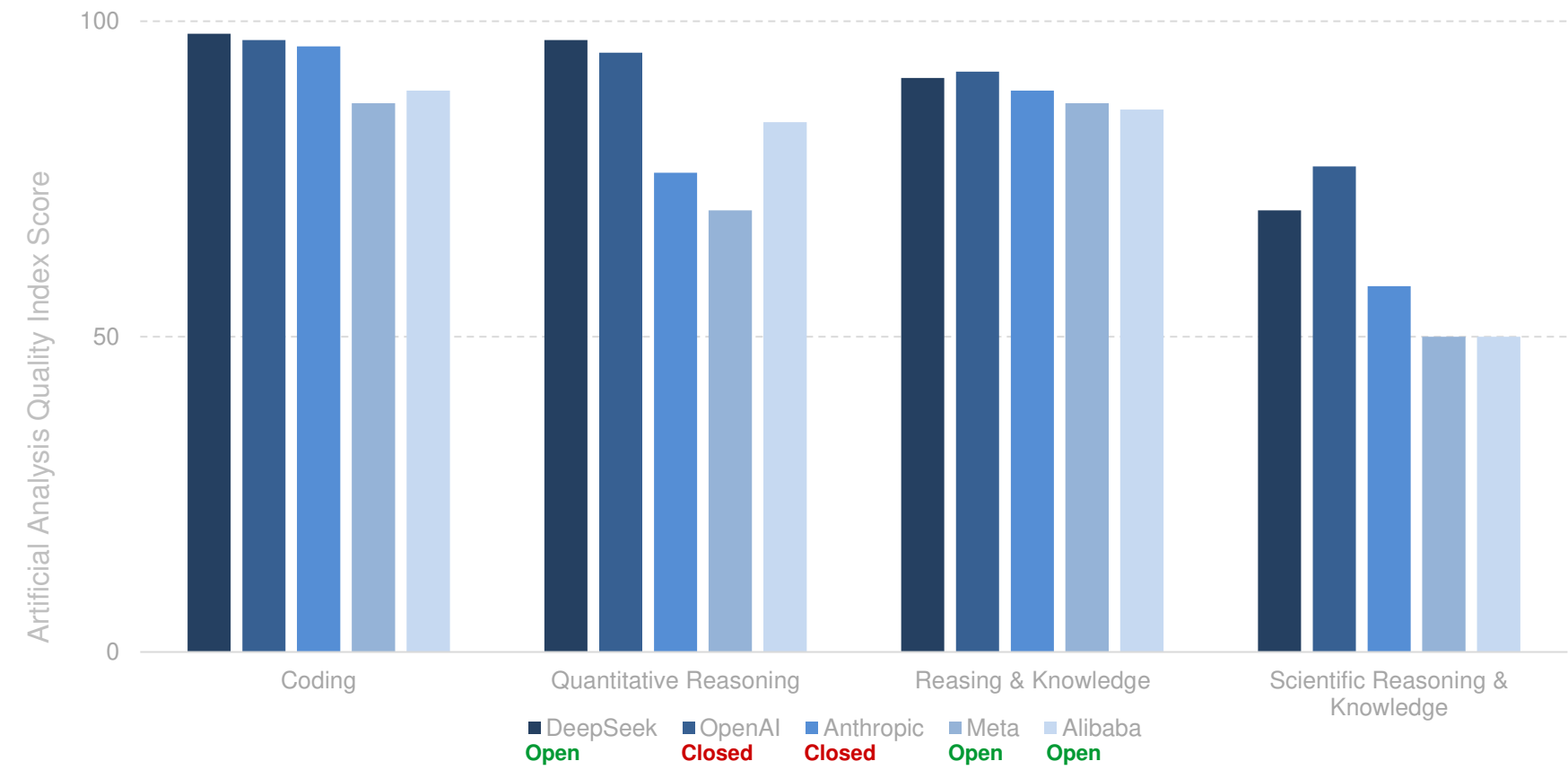


注意：MATH 5 级 pass@1 指的是 AI 模型在 MATH 基准测试上的准确性，MATH 基准测试是高中竞赛级别的数学问题数据集。5 级表示基准测试中最具挑战性的问题。"pass@1" 衡量的是模型是否在其首次尝试中正确解决了问题。来源：Epoch AI (5/25)

...Closed vs. Open-Source Models – Performance = Gap Closing...China Rising, per Artificial Analysis

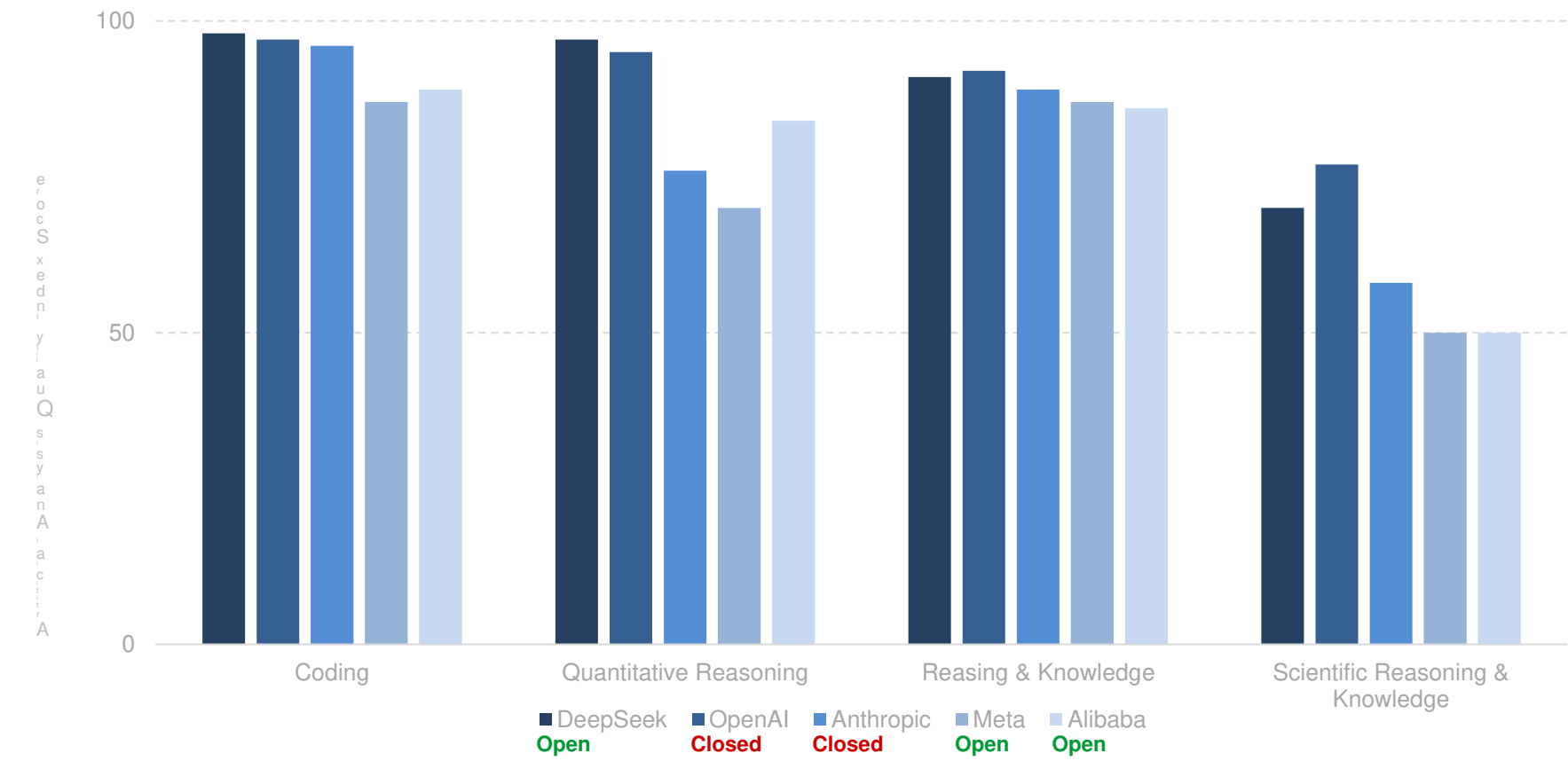
… 封闭 vs. 开放 - 源模型 – 性能 = 差距缩小 … 中国崛起，根据人工智能分析

AI Model Performance by Provider – 1/25, per Artificial Analysis



Note: Scores are out of 100. The models for each company that are measured: for OpenAI, o1; for Alibaba, Qwen 2.5 72B; for Meta, Llama 3.1 405B; for Anthropic, Claude 3.5 Sonnet. The tests used are HumanEval, MATH-500, MMLU and GPQA Diamond. Source: Artificial Analysis via NBC News, 'Why DeepSeek is different, in three charts' (1/25)

按提供商划分的 AI 模型性能 – 1/25, 根据人工智能分析



注：分数为 100 分制。衡量的每家公司的模型为：OpenAI, o1；阿里巴巴, Qwen 2.5 72B；Meta, Llama 3.1 405B；Anthropic, Claude 3.5 Sonnet。使用的测试是 HumanEval、MATH-500、MMLU 和 GPQA Diamond。来源：Artificial Analysis，来自 NBC 新闻，“为什么 DeepSeek 与众不同，见三张图”（1/25）

Rising Performance of Open-Source Models  
+  
Falling Token Costs  
=  
Explosion of Usage by **Developers Using AI**

Rising Performance of Open-Source Models  
+  
Falling Token Costs  
=  
人工智能开发人员使用量激增

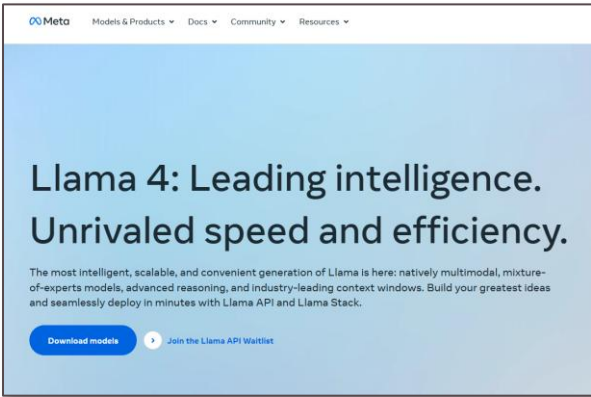
---

<p>Closed-source models – like GPT-4, Claude, or Gemini – have dominated usage among consumers and large enterprises, largely because of their early performance advantage, ease of use, and broader awareness. These models came bundled in clean, productized interfaces and offered reliable outputs with minimal setup.</p> <p>For enterprises, they promise security and ease-of-use for non-technical employees. For consumers, they came with name recognition, fast onboarding, and polished UX. That combination has kept closed models at the center of the AI mainstream.</p> <p>But performance leadership is no longer a given. Open-source models are closing the gap – faster than many expected – and doing so at a fraction of the cost to users. Models like Llama 3 and DeepSeek have demonstrated competitive reasoning, coding, and multilingual abilities, while being fully downloadable, fine-tunable, and deployable on commodity infrastructure.</p> <p>For developers, that matters. Unlike enterprise buyers or end-users, developers care less about polish and more about raw capability, customization, and cost efficiency. And it is developers – more than any other group – who have historically been the leading edge of AI usage. The recent trend appears increasingly clear: more developers are gravitating toward low-cost, high-performance open models, using them to build apps, agents, and pipelines that once required closed APIs.</p> <p>Time will tell if that advantage scales beyond the developer ecosystem. Many open-source tools still lack the brand power, plug-and-play user experience (UX), and managed services that drive adoption among consumers and large organizations. But as the cost-performance ratio of open models continues to improve – and if the infrastructure to support them becomes more turnkey – those advantages could start to spread beyond the developer community.</p>	<p>闭源模型 – 如 GPT-4、 Claude 或 Gemini –ve 在消费者和 ha大型企业中的使用中占据主导地位 es, largely becau由于它们早期的性能优势、易用性和更广泛的认知， 它们被捆绑在简洁、产品化的界面中， These models came b并以最小的成本提供可靠的输出 tup.</p> <p>对于企业来说， 它们承诺为非技术员工提供安全性和易用性。对于消费者来说， 它们具有名称识别、快速入门和精美的用户体验。这种组合使闭源模型始终处于 AI 主流的中心。</p> <p>但性能领先已不再是必然。开源模型正在缩小差距 – 比许多人预期的要快 – 并且以用户成本的一小部分来实现。Llama 3 和 DeepSeek 等模型已经展示了具有竞争力的推理、编码和多语言能力，同时可以完全下载、微调并在通用基础设施上部署。</p> <p>对于开发者来说， 这很重要。与企业买家或最终用户不同，开发者不太关心润色，而更关心原始能力、定制和成本效率。而且开发者 – 比任何其他群体 – 都更能代表人工智能使用的前沿。最近的趋势似乎越来越明显：越来越多的开发者倾向于低成本、高性能的开放模型，使用它们来构建曾经需要封闭 API 的应用程序、代理和管道。</p> <p>时间会证明这种优势是否能扩展到开发者生态系统之外。许多开源工具仍然缺乏品牌影响力、即插即用的用户体验 (UX) 以及推动消费者和大型组织采用的托管服务。但是，随着开放模型的性价比不断提高 – 并且如果支持它们的基础设施变得更加易于使用 – 这些优势可能会开始扩展到开发者社区之外。</p>
---	---

# Developer AI Model Activity = +3.4x Increase in Downloads of Meta Llama in Eight Months

开发者 AI 模型活动 =+3.4x 八个月内 Meta Llama 的下载量增加

Meta Llama – 8/24-4/25, per Meta Platforms



*I predicted that 2025 was going to be the year that open source became the largest type of model that people are developing with, and I think that’s probably going to be the case. That’s kind of how we’re thinking about this overall.*

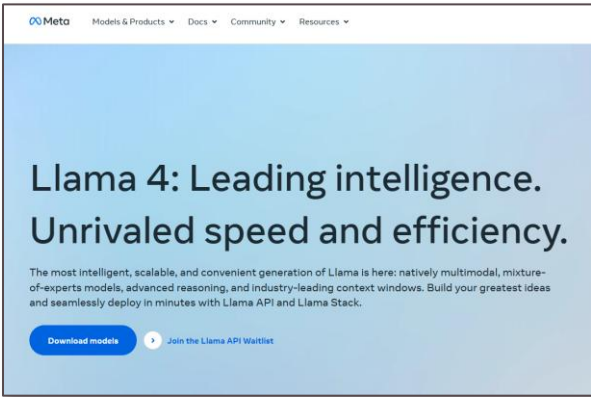
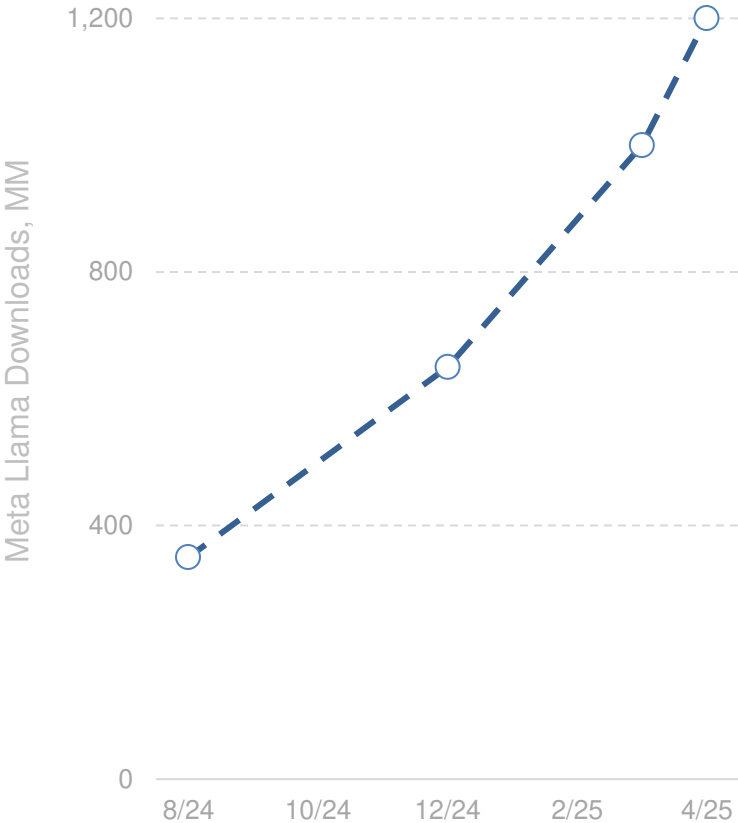
- Meta Platforms CEO Mark Zuckerberg, 5/25

*The groundswell of support for Llama has been awesome. We announced ten weeks ago a billion downloads after the release of Llama 4. In just ten weeks, that number is now 1.2. And if you look at Hugging Face (where the downloads are happening), what’s cool is that most of these are derivatives. We have thousands of developers contributing.*

- Meta Platforms Chief Product Officer Chris Cox, 5/25

Note: 12/24 disclosure counted downloads of Llama and its derivatives. Source: Meta Platforms (8/24, 12/24, 3/25, 4/25), Stratchery podcast (5/25)

Meta Llama Downloads (MM) – 8/24-4/25



我预测 2025 年将成为人们开发的最大类型的开源模型的一年，我认为情况很可能就是这样。这就是我们对整体的看法。

- Meta Platforms CEO Mark Zuckerberg, 5/25

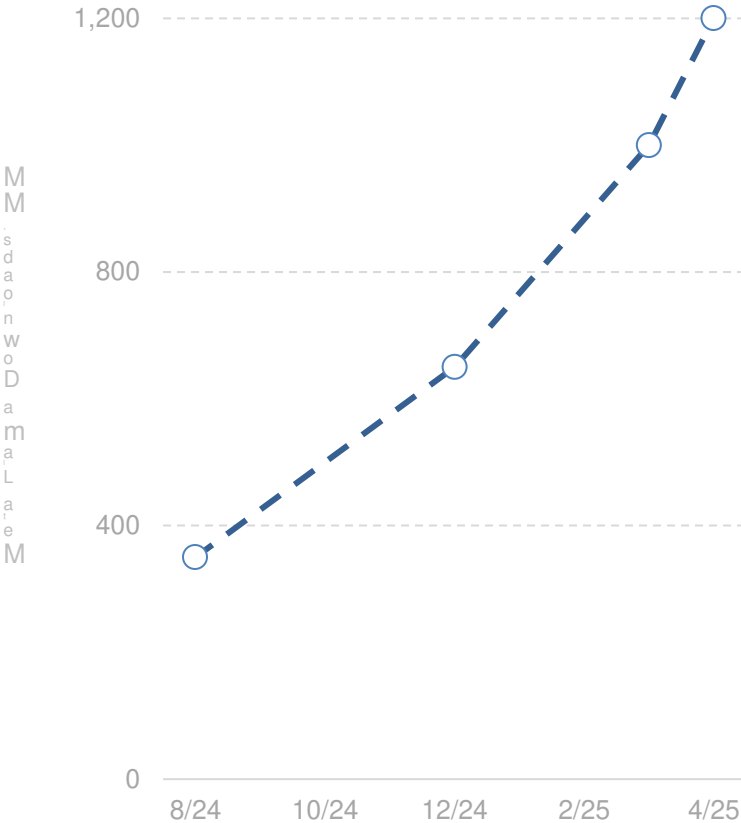
对 Llama 的大力支持令人赞叹。我们在 10 周前宣布，在 Llama 4 发布后下载量达到了 10 亿次。仅仅 10 周后，这个数字现在变成了 12 亿。如果你看看 Hugging Face（下载量发生的地方），最酷的是其中大部分是衍生产品。我们有成千上万的开发者在贡献代码。

- Meta Platforms 首席产品官 Chris Cox, 5/25

Note: 12/24 disclosure 计算了 Llama 及其衍生产品的下载量。来源：Meta Platforms (8/24, 12/24, 3/25, 4/25), Stratchery podcast (5/25)

Meta Llama – 8/24-4/25, 根据 Meta Platforms

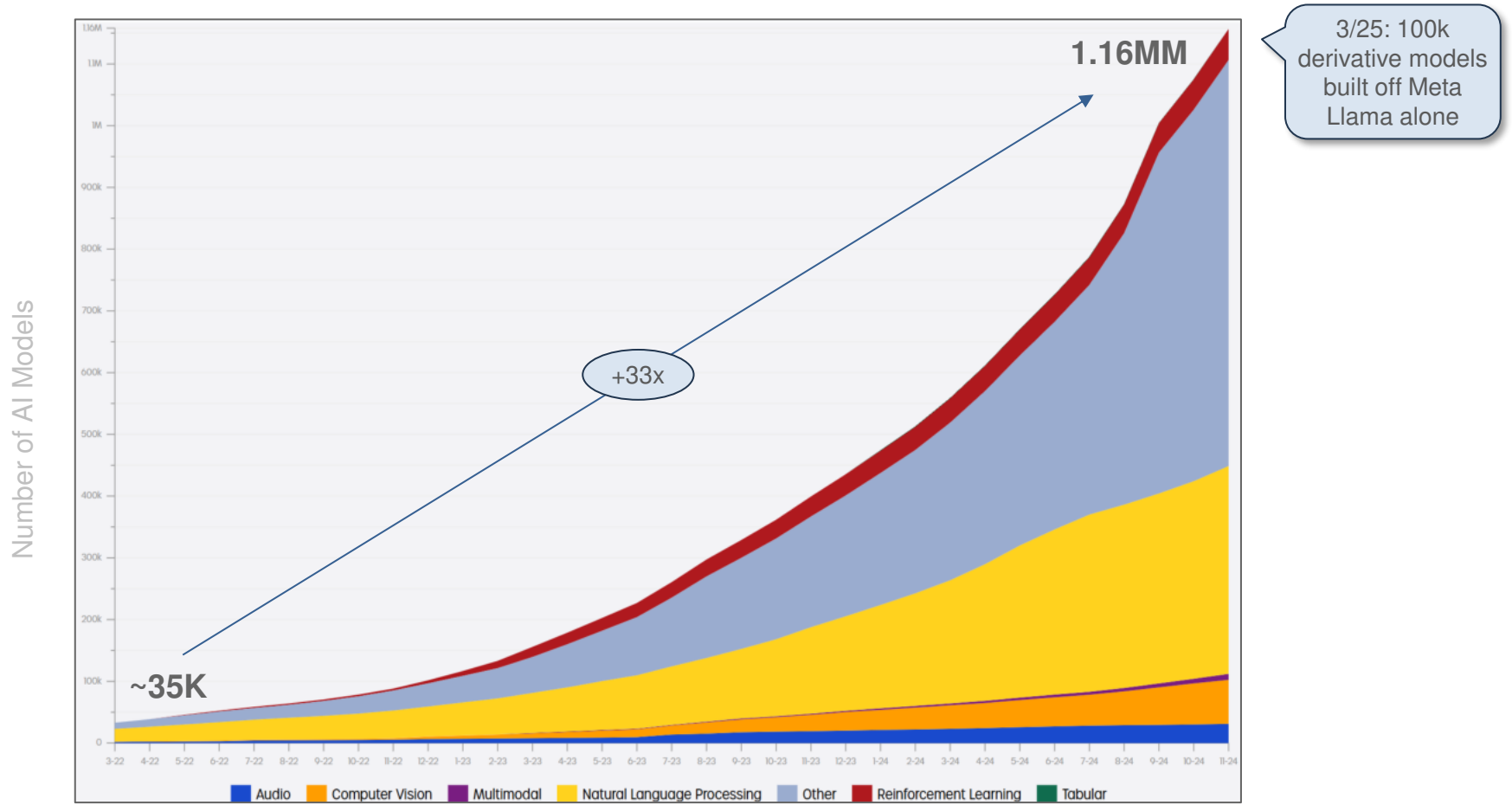
Meta Llama 下载量 ( 百万 ) – 8/24-4/25



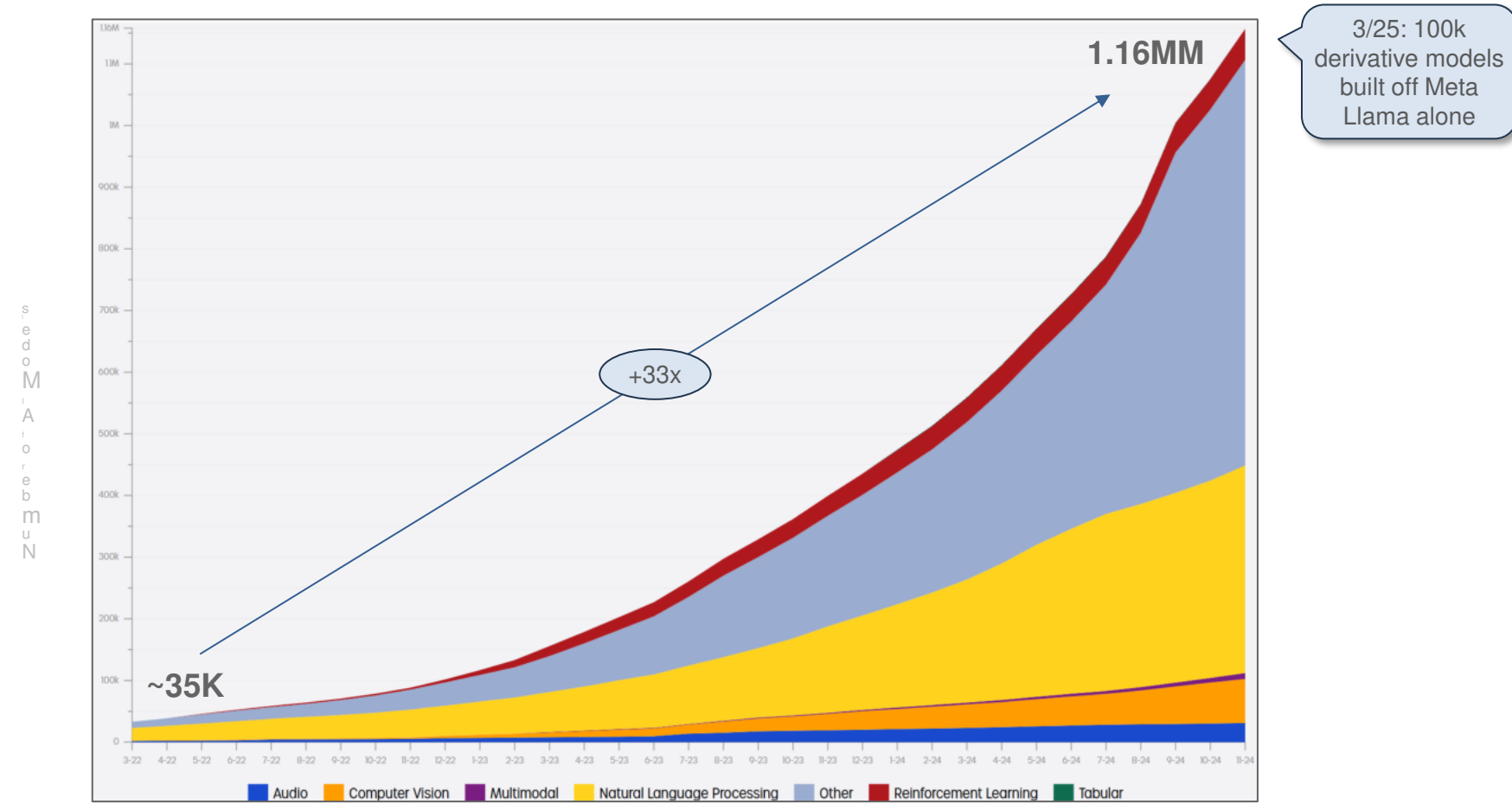
Developer AI Model Activity =  
+33x Increase in AI Models on Hugging Face – 11/24 vs. 3/22

开发者 AI 模型活动 =+33x Hugging Face 上 AI 模型数量增加 –  
11/24 vs. 3/22

AI Models Available from Hugging Face – 3/22-11/24, per Hugging Face



AI Models Available from Hugging Face – 3/22-11/24, per Hugging Face



Note: Hugging Face is an online platform that hosts and shares machine learning models, datasets, and tools – commonly used to access, test, and deploy AI models, including large language models. It has become a central hub for the open-source AI community. May include open-source and closed models. Source: Hugging Face (5/25), Meta (3/25)

注：Hugging Face 是一个在线平台，用于托管和共享机器学习模型、数据集和工具 – 通常用于访问、测试和部署 AI 模型，包括大型语言模型。它已成为开源 AI 社区的中心枢纽。可能包括开源和封闭模型。来源：Hugging Face (5/25)，Meta (3/25)

*AI Monetization Threats*

=

*Rising Competition*

+

*Open-Source Model Momentum*  
**(& China's Rise)**

*AI Monetization Threats*

=

*Rising Competition*

+

*Open-Source Model Momentum*  
**(& China's Rise)**



As noted on page 8, Meta CTO Andrew Bosworth referred to the current state of AI as *our space race and the people we’re discussing, especially China, are highly capable...*

In this context, it is important to remember what the stakes of the Space Race were: proving which political system could innovate faster and win the world’s trust in the process. Coming out on top in the Space Race played a role in enhancing USA’s strategic deterrence and cementing the primacy of western democratic values. The AI ‘space race,’ also has the potential to reshape the world order.

China certainly knows these stakes. Back in 2015, ‘Made in China 2025,’ a new Chinese government initiative to shift the country from low-cost to high-value manufacturing in critical industries, seemed decades away. Fast forward to today, and China has dramatically accelerated its capabilities in these strategic sectors like robotics, electrification, and ‘information technology’ – best expressed by world-class artificial intelligence.

Chinese AI capabilities now underpin nationally strategic areas such as battlefield logistics, target recognition, cyber operations, and autonomous decision-making platforms. In 2025, Chinese state media highlighted the integration of AI into non-combat support functions (e.g., military hospitals), while the Ministry of Science and Technology reinforced its commitment to ‘indigenous innovation’ in strategic technologies.

The implications of Chinese AI supremacy would be profound.

As OpenAI’s Sam Altman noted in a July 2024 *Washington Post* Op-Ed, *If [authoritarian regimes] manage to take the lead on AI, they will force U.S. companies and those of other nations to share user data, leveraging the technology to develop new ways of spying on their own citizens or creating next-generation cyberweapons to use against other countries.*

正如第 8 页所述，Meta 首席技术官安德鲁·博斯沃思将当前人工智能的状态称为我们的太空竞赛，并且我们所讨论的人，尤其是中国，都非常有能力 ……

在这种背景下，重要的是要记住太空竞赛的赌注是什么：证明哪种政治制度能够更快地创新并在这一过程中赢得世界的信任。在太空竞赛中名列前茅，在增强美国战略威慑力量和巩固西方民主价值观的首要地位方面发挥了作用。人工智能 “太空竞赛” 也具有重塑世界秩序的潜力。

中国当然知道这些利害关系。早在 2015 年，中国政府的一项新举措 “中国制造 2025” 旨在推动中国从低成本制造业转向关键行业的高价值制造业，这似乎还需要几十年。快进到今天，中国已大幅提升了其在这些战略领域的能力如机器人、电气化和 “信息技术” —— 最好地体现在世界一流的人工智能上。

中国的人工智能能力现在支撑着国家战略领域，如战场后勤、目标识别、网络行动和自主决策平台。2025 年，中国官方媒体强调了人工智能融入非战斗支援功能（如军事医院），同时科技部重申了其对战略技术 “自主创新” 的承诺。

中国在人工智能领域占据主导地位的影响将是深远的。

正如 OpenAI 的萨姆·奥特曼在 2024 年 7 月《华盛顿邮报》的专栏文章中指出的那样，如果 [专制政权] 设法在人工智能领域取得领先，它们将迫使美国公司和其他国家的公司分享用户数据，利用这项技术开发新的间谍方式来监视自己的公民，或制造下一代网络武器来对付其他国家。

...Meanwhile, alongside AI, broader economic trade tensions between the USA and China continue to escalate, driven by competition for control over strategic technology inputs. China, for now, remains the dominant global supplier of ‘rare earth elements’ – materials essential to advanced electronics, defense systems, and clean energy infrastructure – an imbalance that the USA is working hard to counter. Simultaneously, the USA has prioritized the reshoring of semiconductor manufacturing, supported by the CHIPS and Science Act, and bolstered its partnerships with allied nations (including Japan, South Korea and the Netherlands) to reduce reliance on Chinese supply chains.

Taiwan continues to play a pivotal role in this dynamic. Despite American invention of core semiconductor technology like transistors and EUV lithography, it is Taiwan’s TSMC – the world’s most advanced semiconductor foundry – that drives global semiconductor production and is therefore central to both countries’ strategic calculations.

It has taken a long time for the USA to wake up, but after two decades of inaction, both political parties are calling loudly for change. While each has taken a different approach (export controls in the Biden administration, economic nationalism and reshoring in the Trump administration), the move towards treating cutting-edge technology development as a core part of the national interest is a welcome adjustment.

As Senators John Cornyn and Mark Warner noted in 2020 regarding semiconductors, *America’s innovation in semiconductors undergirds our entire innovation economy...unfortunately, our complacency has allowed our competitors – including adversaries – to catch up.*

However, despite these measures, American intellectual property remains at risk; per OpenAI, *We know PRC (China) based companies – and others – are constantly trying to distill the models of leading US AI companies...it is critically important that we are working closely with the US government to best protect the most capable models from efforts by adversaries and competitors to take US technology.*

What is clear, however, is that the American tone about Chinese technology has morphed since the early 2000s enthusiasm around China’s entry into the World Trade Organization (WTO). AI, semiconductors, and critical minerals, and technology developments in general, are no longer viewed solely as economic or technology assets – they represent strategic levers of national resilience and geopolitical power, core to both the USA and China.

…与此同时，除了人工智能之外，美国和中国之间更广泛的经济贸易紧张局势持续升级，其原因是争夺对战略技术投入的控制权。目前，中国仍然是全球主要的 “ 稀土元素 ” 供应商 —— 这些材料对先进电子产品、国防系统和清洁能源基础设施至关重要 —— 美国正在努力改变这种不平衡的局面。与此同时，美国已将半导体制造业的回流列为优先事项，并得到了《芯片与科学法案》的支持，并加强了与盟国的伙伴关系（包括日本、韩国和荷兰），以减少对中国供应链的依赖。

台湾继续在这种动态中发挥关键作用。尽管美国发明了核心半导体技术，如晶体管和 EUV 光刻技术，但正是台湾的台积电（TSMC）——世界上最先进的半导体代工厂——推动着全球半导体生产，因此对两国的战略考量都至关重要。

美国花了很长时间才醒悟过来，但在经历了二十年的无所作为之后，两党都在大声疾呼要求变革。虽然双方采取了不同的方法（拜登政府的出口管制，特朗普政府的经济民族主义和回流），但将尖端技术发展视为国家利益核心部分的举措是一种可喜的调整。

正如参议员 John Cornyn 和 Mark Warner 在 2020 年关于半导体问题上指出的那样，美国在半导体领域的创新支撑着我们整个创新经济 …… 不幸的是，我们的自满情绪让我们的竞争对手 —— 包括对手 —— 赶了上来。

然而，尽管采取了这些措施，美国的知识产权仍然面临风险；根据 OpenAI 的说法，我们知道总部位于中国（PRC）的公司 —— 和其他公司 —— 一直在试图提炼领先的美国人工智能公司的模型 …… 至关重要的是，我们正在与美国政府密切合作，以最好地保护最有能力的模型，使其免受对手和竞争对手获取美国技术的企图。

然而，显而易见的是，自从 2000 年代初期对中国加入世界贸易组织（WTO）的热情以来，美国对中国技术的论调已经发生了变化。人工智能、半导体和关键矿产，以及一般的技术发展，不再仅仅被视为经济或技术资产 —— 它们代表着国家韧性和地缘政治力量的战略杠杆，对美国和中国都至关重要。

Public Market Capitalization Leader  
Tells of Last Thirty Years =

Extraordinary USA Momentum...  
China Rising

上市公司市值领导者讲述过去三十年 =

非凡的美国势头 …… 中国崛起

Global Public Market Capitalization Leaders – May, 2025 = 83% (25 of 30) USA-Based...

Global Public Companies Ranked By Market Capitalization – 5/15/25, per Capital IQ

Rank 2025	Company	HQ Country	Sector	Market Cap (\$B)
1	Microsoft	USA	Software / AI	\$3,368B
2	NVIDIA	USA	Semis / AI	3,288
3	Apple	USA	Hardware / AI	3,158
4	Amazon	USA	Internet / AI	2,178
5	Alphabet (Google)	USA	Internet / AI	1,997
6	Saudi Aramco	Saudi Arabia	Energy	1,686
7	Meta Platforms (Facebook)	USA	Internet / AI	1,619
8	Tesla	USA	Auto / AI	1,104
9	Broadcom	USA	Semis / AI	1,094
10	Berkshire Hathaway	USA	Finance	1,093
11	TSMC	Taiwan	Semis / AI	856
12	Walmart	USA	Consumer Products	771
13	JP Morgan Chase	USA	Finance	743
14	Visa	USA	Finance	678
15	Eli Lilly	USA	Healthcare	658
16	Tencent	China	Software / AI	591
17	Mastercard	USA	Finance	529
18	Netflix	USA	Internet / AI	501
19	Exxon Mobil	USA	Energy	468
20	Costco Wholesale	USA	Consumer Products	448
21	Oracle	USA	Hardware / AI	447
22	Procter & Gamble	USA	Consumer Products	381
23	Home Depot	USA	Consumer Products	376
24	Johnson & Johnson	USA	Consumer Products	360
25	SAP	Germany	Software / AI	343
26	Bank of America	USA	Finance	334
27	ICBC	China	Finance	330
28	AbbVie	USA	Healthcare	321
29	Coca-Cola	USA	Consumer Products	308
30	Palantir	USA	Software / AI	302

Source: Capital IQ (as of 5/15/25)

全球公开市场资本总额领导者 – 2025 = 月， 83% (30 家中的 25 家 ) 美国公司 …

全球上市公司按市值排名 – 5/15/25, 根据 Capital IQ

Rank 2025	公司	总部国家	行业	市值 (\$B)
1	Microsoft	USA	软件 / AI	\$3,368B
2	NVIDIA	USA	半导体 / AI	3,288
3	Apple	USA	硬件 / AI	3,158
4	Amazon	USA	互联网 / AI	2,178
5	Alphabet (Google)	USA	互联网 / AI	1,997
6	Saudi Aramco	沙特阿拉伯	能源	1,686
7	Meta Platforms (Facebook)	USA	互联网 / AI	1,619
8	Tesla	USA	汽车 / AI	1,104
9	Broadcom	USA	半导体 / AI	1,094
10	伯克希尔·哈撒韦	USA	金融	1,093
11	TSMC	Taiwan	半导体 / A I	856
12	沃尔玛	USA	Consumer Products	771
13	摩根大通	USA	金融	743
14	Visa	USA	金融	678
15	礼来公司	USA	医疗保健	658
16	Tencent	中国	软件 / AI	591
17	万事达卡	USA	金融	529
18	Netflix	USA	互联网 / AI	501
19	埃克森美孚	USA	能源	468
20	好市多	USA	消费品	448
21	Oracle	USA	硬件 / AI	447
22	Procter & Gamble	USA	消费品	381
23	家得宝	USA	消费品	376
24	强生	USA	消费品	360
25	SAP	Germany	软件 /AI	343
26	美国银行	USA	金融	334
27	ICBC	中国	金融	330
28	AbbVie	USA	医疗保健	321
29	Coca-Cola	USA	消费品类 roducts	308
30	Palantir	USA	Software / AI	302

Source: Capital IQ (as of 5/15/25)

...Global Public Market Capitalization Leaders – December, 1995 = 53% (16 of 30) USA-Based

Global Public Companies Ranked By Market Capitalization – 12/31/95, per Bloomberg

Rank 1995	Company	HQ Country	Sector	Market Cap (\$B)
1	Nippon Telegraph	Japan	Telco	\$128B
2	General Electric	USA	Industrials	120
3	AT&T	USA	Telco	103
4	Exxon	USA	Energy	100
5	Coca-Cola	USA	Consumer Products	94
6	Merck	USA	Healthcare	81
7	Toyota	Japan	Automotive	79
8	Roche	Switzerland	Healthcare	78
9	Altria	USA	Consumer Products	75
10	Industrial Bank of Japan	Japan	Finance	71
11	MUFG Bank	Japan	Finance	68
12	Sumimoto Mitsui	Japan	Finance	66
13	Fuji Bank	Japan	Finance	64
14	Dai-Ichi Kangyo Bank	Japan	Finance	61
15	UFJ Bank	Japan	Finance	59
16	Novartis	Switzerland	Healthcare	57
17	Procter & Gamble	USA	Consumer Products	57
18	Johnson & Johnson	USA	Consumer Products	55
19	Microsoft	USA	Software	52
20	Walmart	USA	Consumer Products	51
21	IBM	USA	Hardware / Software	51
22	DirecTV	USA	Media	49
23	Intel	USA	Hardware	47
24	BP	United Kingdom	Energy	46
25	Nestle	Switzerland	Consumer Products	45
26	Mobil	USA	Energy	44
27	PepsiCo	USA	Consumer Products	44
28	AIG	USA	Finance	44
29	Shell	United Kingdom	Energy	44
30	Sakura Bank	Japan	Finance	43

Source: Bloomberg (as of 5/15/25)

… 全球上市公司市值领导者 – December, 1995 =53% (16 of 30) USA-Based

按市值排名的全球上市公司 – 12/31/95，根据彭博数据

排名 1995	公司	总部国家	行业	市值 (\$B)
1	Nippon Telegraph	日本	电信	\$128B
2	通用电气	USA	工业	120
3	AT&T	USA	电信	103
4	Exxon	USA	能源	100
5	Coca-Cola	USA	消费品	94
6	默克	USA	医疗保健	81
7	Toyota	日本	汽车	79
8	罗氏	瑞士	医疗保健	78
9	奥驰亚	USA	消费品	75
10	行业 rial Bank of Japan	日本	金融	71
11	三菱日联银行	日本	金融	68
12	住友三井	日本	金融	66
13	富士银行	日本	金融	64
14	第一劝业银行	日本	金融	61
15	UFJ 银行	日本	金融	59
16	Novartis	瑞士	医疗保健	57
17	Procter & Gamble	USA	消费品	57
18	强生	USA	消费品	55
19	Microsoft	USA	软件	52
20	沃尔玛	USA	消费品	51
21	IBM	USA	硬件 / 软件	51
22	DirecTV	USA	媒体	49
23	英特尔	USA	硬件	47
24	BP	联合 Kingdom	能源	46
25	Nestle	瑞士	消费品 er Products	45
26	Mobil	USA	能源	44
27	PepsiCo	USA	消费品	44
28	AIG	USA	金融	44
29	壳牌	联合 王国	能源	44
30	Sakura Bank	Japan	Finance	43

Source: Bloomberg (as of 5/15/25)

Over the past thirty years (1995 to 2025), just six companies remained on the top 30 most highly valued **publicly traded global companies** – Microsoft / Walmart / Exxon Mobil / Procter & Gamble / Johnson & Johnson / Coca-Cola.

New entrants are NVIDIA / Apple / Amazon / Alphabet (Google) / Saudi Aramco / Meta Platforms (Facebook) / Tesla / Broadcom / Berkshire Hathaway / TSMC / JP Morgan Chase / Visa / Eli Lilly / Tencent / Mastercard / Netflix / Costco Wholesale / Oracle / Home Depot / SAP / Bank of America / ICBC / AbbVie / Palantir.

**In 1995, USA had 53% (16 of 30) of the most valuable companies and 83% (25 of 30) in 2025.**  
Japan came next with 9, now 0.  
Switzerland followed with 3, now 0. UK had 2, now 0.

In 2025, new geographic entrants include China with 2 and Saudi Arabia / Taiwan / Germany with 1 each.

在过去的三十年（1995 年至 2025 年）里，只有六家公司始终位居全球 30 家市值最高的上市公司之列 –Microsoft / Walmart / Exxon Mobil / Procter & Gamble / Johnson & Johnson / Coca-Cola。

新的入选者包括 NVIDIA / Apple / Amazon / Alphabet (Google) / Saudi Aramco / Meta Platforms (Facebook) / Tesla / Broadcom / Berkshire Hathaway / TSMC / JP Morgan Chase / Visa / Eli Lilly / Tencent / Mastercard / Netflix / Costco Wholesale / Oracle / Home Depot / SAP / Bank of America / ICBC / AbbVie / Palantir。

1995 年，美国占最有价值公司的 53% （30 家中的 16 家），2025 年占 83% （30 家中的 25 家）。其次是日本，当时占 9 家，现在为 0。瑞士紧随其后，当时占 3 家，现在为 0。英国当时占 2 家，现在为 0。

2025 年，新的地区入选者包括中国，有 2 家，沙特阿拉伯 / 台湾 / 德国各有 1 家。



Global Public Technology Market Cap Leaders – May, 2025 = 70% (21 of 30) USA-Based...

Global Technology Companies Ranked By Market Capitalization – 5/15/25, per Capital IQ

Rank 2025	Company	HQ Country	Sector	Market Cap (\$B)
1	Microsoft	USA	Software / AI	\$3,368B
2	NVIDIA	USA	Semis / AI	3,288
3	Apple	USA	Hardware / AI	3,158
4	Amazon	USA	Internet / AI	2,178
5	Alphabet (Google)	USA	Internet / AI	1,997
6	Meta Platforms (Facebook)	USA	Internet / AI	1,619
7	Tesla	USA	Auto / AI	1,104
8	Broadcom	USA	Semis / AI	1,094
9	TSMC	Taiwan	Semis / AI	856
10	Tencent	China	Software / AI	591
11	Netflix	USA	Internet / AI	501
12	Oracle	USA	Hardware / AI	447
13	SAP	Germany	Software / AI	343
14	Palantir	USA	Software / AI	302
15	ASML	Netherlands	Semis / AI	300
16	Alibaba	China	Internet / AI	281
17	Salesforce	USA	Software / AI	279
18	T-Mobile	USA	Telco	273
19	Samsung	S. Korea	Hardware / AI	268
20	Cisco	USA	Semis / AI	256
21	IBM	USA	Hardware / AI	243
22	China Mobile	China	Telco	241
23	Reliance	India	Telco	216
24	ServiceNow	USA	Software / AI	214
25	Intuitive Surgical	USA	Health Tech	201
26	AT&T	USA	Telco	197
27	Siemens	Germany	Hardware / AI	194
28	Uber	USA	Internet / AI	189
29	AMD	USA	Semis / AI	186
30	Intuit	USA	Software / AI	185

Source: Capital IQ (as of 5/15/25)

全球公共科技市值领导者 –2025 = 年 5 月， 70% （ 30 家中的 21 家）总部位于美国 ……

按市值排名的全球科技公司 –5/15/25， 数据来源： Capital IQ

排名 2025	公司	总部国家	行业	市值 (\$B)
1	Microsoft	USA	软件 / AI	\$3,368B
2	NVIDIA	USA	半导体 /AI	3,288
3	Apple	USA	硬件 /AI	3,158
4	Amazon	USA	互联网 / AI	2,178
5	Alphabet (Google)	USA	互联网 / AI	1,997
6	Meta Platforms (Facebook)	USA	互联网 / AI	1,619
7	特斯拉	USA	汽车 / AI	1,104
8	博通	USA	半导体 / AI	1,094
9	TSMC	台湾	半导体 /AI	856
10	Tencent	中国	软件 /AI	591
11	Netflix	USA	互联网 / AI	501
12	Oracle	USA	硬件 / AI	447
13	SAP	Germany	软件 /AI	343
14	Palantir	USA	软件 / AI	302
15	ASML	荷兰	半导体 /AI	300
16	阿里巴巴	中国	互联网 /AI	281
17	Salesforce	USA	软件 / AI	279
18	T-Mobile	USA	电信公司	273
19	Samsung	S. Korea	Hardware / AI	268
20	Cisco	USA	半导体 /AI	256
21	IBM	USA	硬件 / AI	243
22	中国移动	中国	电信公司	241
23	Reliance	印度	电信公司	216
24	ServiceNow	USA	软件 / AI	214
25	Intuitive Surgical	USA	健康科技	201
26	AT&T	USA	电信	197
27	西门子 ns	Germany	硬件 / AI	194
28	Uber	USA	互联网 /AI	189
29	AMD	USA	半导体 /AI	186
30	Intuit	USA	Software / AI	185

Source: Capital IQ (as of 5/15/25)



...Global Public Technology Market Cap Leaders – December, 1995 = 53% (16 of 30) USA-Based

Global Technology Companies Ranked By Market Capitalization – 12/31/95, per Bloomberg

Rank 1995	Company	HQ Country	Sector	Market Cap (\$B)
1	Nippon Telegraph	Japan	Telco	\$128B
2	AT&T	USA	Telco	103
3	Microsoft	USA	Software	52
4	IBM	USA	Hardware / Software	51
5	Intel	USA	Hardware	47
6	BellSouth	USA	Telco	43
7	HP	USA	Hardware	43
8	GTE	USA	Telco	42
9	BT	United Kingdom	Telco	34
10	Panasonic	Japan	Hardware	34
11	SingTel	Singapore	Telco	34
12	Motorola	USA	Hardware	34
13	Hitachi	Japan	Hardware	33
14	Verizon	USA	Telco	29
15	Toshiba	Japan	Hardware	26
16	Peraton	USA	Software / Hardware	25
17	Nynex	USA	Telco	24
18	Sony	Japan	Hardware	22
19	Cisco	USA	Hardware	21
20	Fujitsu	Japan	Hardware	20
21	PCCW	Hong Kong	Telco	20
22	NEC	Japan	Software	19
23	Oracle	USA	Hardware	18
24	MCI	USA	Telco	18
25	Sharp	Japan	Hardware	18
26	TelMex	Mexico	Telco	17
27	KDDI	Japan	Telco	17
28	US West	USA	Telco	17
29	Cable & Wireless	USA	Telco	16
30	Telekom Malaysia	Malaysia	Telco	16

Source: Bloomberg (as of 5/15/25)

… 全球上市科技公司市值领导者 –1995 =月， 53% （ 30 家中的 16 家） 位于美国

按市值排名的全球科技公司 –1995 年 12 月 31 日， 数据来源： 彭博

Rank 1995	公司	总部国家	行业	市值 (\$B)
1	Nippon Telegraph	日本	电信	\$128B
2	AT&T	USA	电信	103
3	Microsoft	USA	软件	52
4	IBM	USA	硬件 / 软件	51
5	Intel	USA	硬件	47
6	BellSouth	USA	电信	43
7	HP	USA	硬件	43
8	GTE	USA	电信	42
9	BT	英国	电信公司	34
10	Panasonic	日本	硬件	34
11	SingTel	新加坡	电信公司	34
12	摩托罗拉	USA	硬件	34
13	日立	日本	硬件	33
14	Verizon	USA	电信公司	29
15	东芝 (Toshiba)	日本	硬件	26
16	Peraton	USA	软件 / Hardware	25
17	Nynex	USA	电信公司	24
18	Sony	日本	硬件	22
19	Cisco	USA	硬件	21
20	Fujitsu	日本	硬件	20
21	PCCW	香港 K ong	电信	20
22	NEC	日本	软件	19
23	Oracle	USA	硬件	18
24	MCI	USA	电信	18
25	夏普	日本	硬件	18
26	TelMex	墨西哥	电信公司	17
27	KDDI	日本	电信	17
28	美国西部	USA	电信	17
29	Cable & Wireless	USA	Telco	16
30	Telekom Malaysia	Malaysia	Telco	16

Source: Bloomberg (as of 5/15/25)

Over the past thirty years (1995 to 2025), just five companies remained on the top 30 most highly valued **publicly traded global technology companies** – Microsoft / Oracle / Cisco / IBM / AT&T.

New entrants are NVIDIA / Apple / Amazon / Alphabet (Google) / Meta Platforms (Facebook) / Tesla / Broadcom / TSMC / Tencent / Netflix / SAP / Palantir / ASML / Alibaba / Salesforce / T-Mobile / Samsung / China Mobile / Reliance / ServiceNow / Intuitive Surgical / Siemens / Uber / AMD / Intuit.

**In 1995, USA had 53% (16 of 30) of the most valuable tech companies and 70% (21 of 30) in 2025.**

In 1995, Japan had 30% (9 of 30) of the top tech companies and 0 in 2025. UK / Singapore / Hong Kong / Mexico / Malaysia had 1, now 0.

In 2025, new geographic entrants include China with 3, Germany with 2, Taiwan with 1, Netherlands with 1, South Korea with 1 & India with 1.

Note that while Taiwan has only one company on the list – TSMC – the company produces 80%-90% of the world’s most advanced semiconductors and 62%+ of global semiconductors as of Q2:24, per The Center for Strategic & International Studies & Counterpoint Research.

**It’s stunning how much can change in a generation... the emergence of internet connectivity was foundational to most of the new adds. The emergence of AI will have the same type of effect over the next three decades, but likely faster.**

在过去的三十年（1995 年至 2025 年）里，只有五家公司仍然位居全球最具价值的 30 家上市科技公司之列 –Microsoft / Oracle / Cisco / IBM / AT&T。

新进者包括 NVIDIA / Apple / Amazon / Alphabet (Google) / Meta Platforms (Facebook) / Tesla / Broadcom / TSMC / Tencent / Netflix / SAP / Palantir / ASML / Alibaba / Salesforce / T-Mobile / Samsung / China Mobile / Reliance / ServiceNow / Intuitive Surgical / Siemens / Uber / AMD / Intuit。

1995 年，美国拥有 53% （30 家中的 16 家）最有价值的科技公司，2025 年为 70% （30 家中的 21 家）。1995 年，日本拥有 30% （30 家中的 9 家）的顶级科技公司，而 2025 年为 0。英国 / 新加坡 / 香港 / 墨西哥 / 马来西亚曾经有 1 家，现在是 0。

2025 年，新的地理区域进入者包括中国（3 家）、德国（2 家）、台湾（1 家）、荷兰（1 家）、韩国（1 家）和印度（1 家）。

请注意，虽然台湾在该榜单上只有一家公司 – 台积电 TSMC–，但根据战略与国际研究中心和 Counterpoint Research 的数据，截至 24 年第二季度，该公司生产了全球 80%-90% 的最先进的半导体和 62%+ 的全球半导体。

**一代人的时间可以改变这么多，真是令人震惊 ……** 互联网连接的出现是大多数新增内容的基础。人工智能的出现将在未来三十年内产生相同类型的影响，但可能会更快。

Source: Center for Strategic & International Studies, 'A Strategy for The United States to Regain its Position in Semiconductor Manufacturing' (2/24); Counterpoint Research, 'Global Semiconductor Foundry Market Share: Quarterly' (3/25)

来源：战略与国际研究中心，《美国重获半导体领域地位的战略》（2/24）；Counterpoint Research，《全球半导体代工厂市场份额：季度》（3/25）

USA vs. China in Technology =

China's AI Response Time  
Significantly Faster vs. Internet 1995

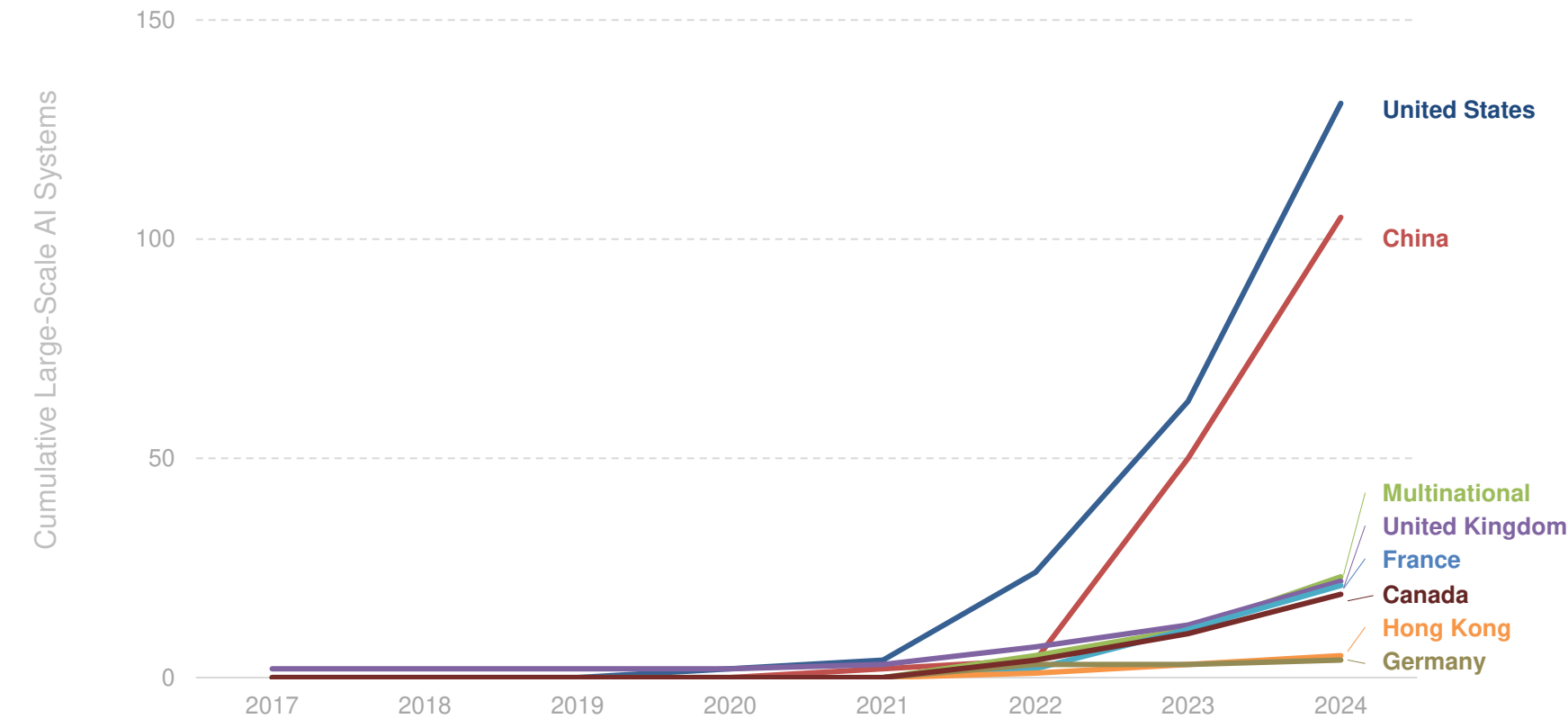
USA vs. China in Technology =

China's AI Response Time  
Significantly Faster vs. Internet 1995

# AI Large Language Model (LLM) Leadership = USA & China Outpacing Rest of World (RoW), per Epoch AI

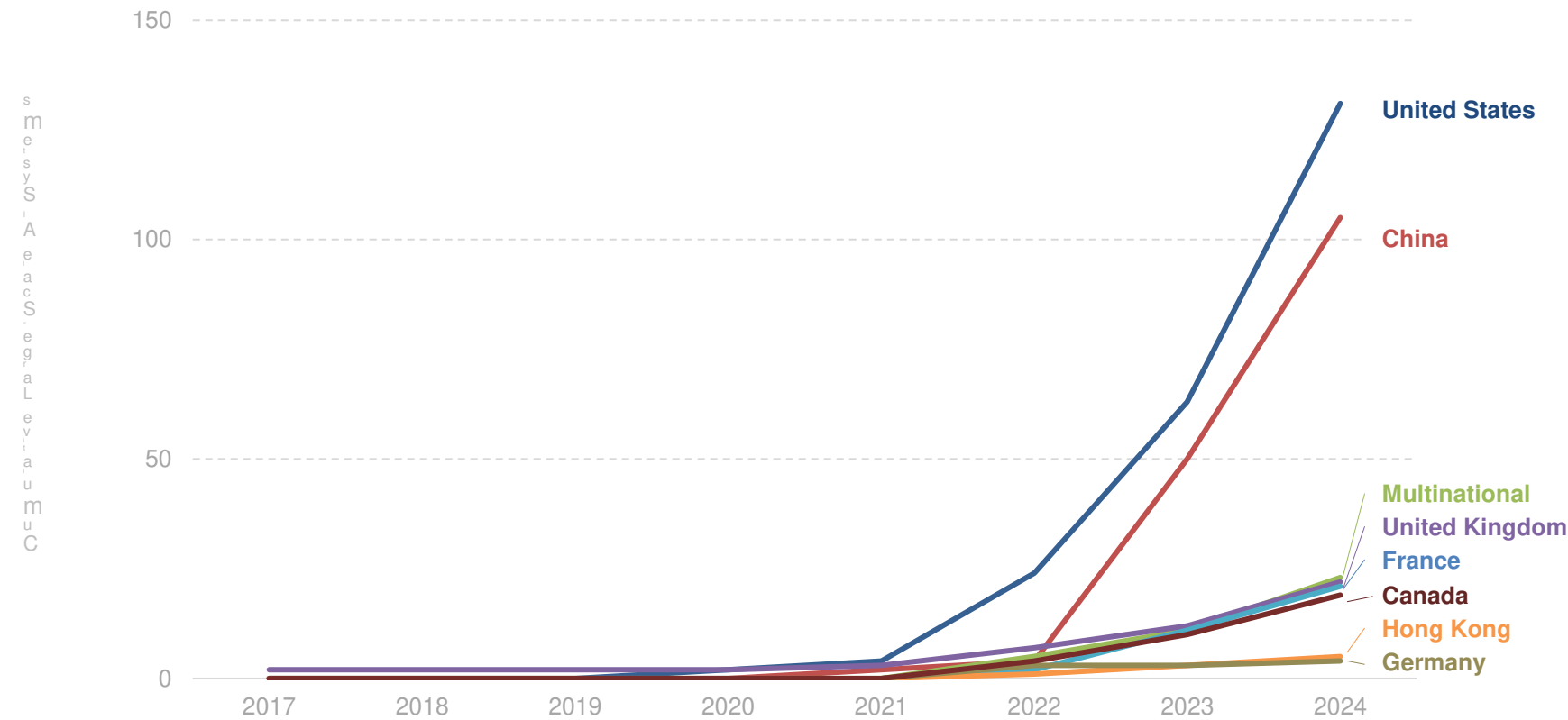
# AI 大型语言模型（LLM）领导地位 = 美国和中国超越世界其他地区（RoW），数据来源：Epoch AI

Cumulative Large-Scale AI Systems by Country\* – 2017-2024, per Epoch AI



\*Hong Kong is a Special Administrative Region (SAR) of China, not an independent country. Note: Epoch AI defines AI models as 'large-scale' when their training compute is confirmed to exceed 10<sup>23</sup> floating-point operations. Source: Epoch AI via Our World In Data (5/25)

按国家 / 地区划分的累积大规模 AI 系统 \* – 2017-2024，数据来源：Epoch AI



\* 香港是中国的特别行政区（SAR），不是一个独立的国家。注意：Epoch AI 将 AI 模型定义为 'large-scale'，当确认它们的训练计算量超过 10<sup>23</sup> 浮点运算时。来源：Epoch AI，通过 Our World In Data (5/25)

WorldBusinessMarketsSustainabilityLegalBreakingviewsTechnologyInvestigationsMore

# DeepSeek narrows China-US AI gap to three months, 01.AI founder Lee Kai-fu says

By Liam Mo and Kane Wu

March 25, 2025 12:08 AM PDT · Updated 20 hours ago

Aa

The Deepseek logo is seen in this illustration taken Jan. 27, 2025. REUTERS/Dado Ruvic/llustration/File Photo [Purchase Licensing Rights](#)

BEIJING/HONG KONG, March 25 (Reuters) - China has narrowed the AI development gap with the United States to just three months in some areas, because firms such as DeepSeek have worked out how to use chips and apply algorithms more efficiently, the CEO of Chinese startup 01.AI Lee Kai-fu said.

Lee, a prominent figure in the global artificial intelligence space and a former head of Google China, told Reuters that startup DeepSeek revealed that China had pulled ahead in areas such as infrastructure software engineering

*We believe that as the economy develops, China should gradually become a contributor instead of freeriding. In the past 30+ years of the IT wave, we basically didn’t participate in real technological innovation. We’re used to Moore’s Law falling out of the sky, lying at home waiting 18 months for better hardware and software to emerge. That’s how the Scaling Law is being treated...*

*What we see is that Chinese AI can’t be in the position of following forever. We often say that there is a gap of one or two years between Chinese AI and the United States, but the real gap is the difference between originality and imitation. If this doesn’t change, China will always be only a follower – so some exploration is inescapable.*

- DeepSeek CEO Liang Wenfang, 11/24

WorldBusinessMarketsSustainabilityLegalBreakingviewsTechnologyInvestigationsMore

# DeepSeek narrows China-US AI gap to three months, 01.AI founder Lee Kai-fu says

By Liam Mo and Kane Wu

March 25, 2025 12:08 AM PDT · Updated 20 hours ago

Aa

The Deepseek logo is seen in this illustration taken Jan. 27, 2025. REUTERS/Dado Ruvic/llustration/File Photo [Purchase Licensing Rights](#)

BEIJING/HONG KONG, March 25 (Reuters) - China has narrowed the AI development gap with the United States to just three months in some areas, because firms such as DeepSeek have worked out how to use chips and apply algorithms more efficiently, the CEO of Chinese startup 01.AI Lee Kai-fu said.

Lee, a prominent figure in the global artificial intelligence space and a former head of Google China, told Reuters that startup DeepSeek revealed that China had pulled ahead in areas such as infrastructure software engineering

我们认为，随着经济的发展，中国应该逐渐成为贡献者，而不是搭便车的人。在过去的 30+ 几年里的 IT 浪潮中，我们基本上没有参与真正的技术创新。我们已经习惯了摩尔定律从天而降，在家等待 18 个月，更好的硬件和软件就会出现。规模法则就是这样被对待的 ……

我们看到的是，中国人工智能不能永远处于追随的地位。我们经常说中国人工智能与美国之间存在一两年的差距，但真正的差距是原创与模仿之间的差异。如果这种情况不改变，中国将永远只是一个追随者 —— 因此，一些探索是不可避免的。

- DeepSeek CEO Liang Wenfang, 11/24



Mashable

Meet Alibaba's Qwen 2.5, an AI model claiming to beat both DeepSeek and OpenAI's ChatGPT

Mashable

Wed, January 29, 2025 at 9:11 AM PST · 1 min read




Photo Illustration by Thomas Fuller/SOPA Images/LightRocket via Getty Images

Alibaba, the Chinese tech giant, released a new version of its AI model and made big claims — notably that it outperforms OpenAI's ChatGPT and the [newly ascending DeepSeek](#).

*Qwen2.5-Max outperforms DeepSeek V3 in benchmarks such as Arena-Hard, LiveBench, LiveCodeBench, and GPQA-Diamond, while also demonstrating competitive results in other assessments, including MMLU-Pro.*

*Our base models have demonstrated significant advantages across most benchmarks, and we are optimistic that advancements in post-training techniques will elevate the next version of Qwen2.5-Max to new heights.*

*The scaling of data and model size not only showcases advancements in model intelligence but also reflects our unwavering commitment to pioneering research. We are dedicated to enhancing the thinking and reasoning capabilities of large language models through the innovative application of scaled reinforcement learning.*

**- Alibaba Qwen 2.5 Press Release, 1/25**

Mashable

Meet Alibaba's Qwen 2.5, an AI model claiming to beat both DeepSeek and OpenAI's ChatGPT

Mashable

Wed, January 29, 2025 at 9:11 AM PST · 1 min read




Photo Illustration by Thomas Fuller/SOPA Images/LightRocket via Getty Images

Alibaba, the Chinese tech giant, released a new version of its AI model and made big claims — notably that it outperforms OpenAI's ChatGPT and the [newly ascending DeepSeek](#).

Qwen2.5-Max 在 Arena-Hard、LiveBench、LiveCodeBench 和 GPQA-Diamond 等基准测试中优于 DeepSeek V3，同时在包括 MMLU-Pro 在内的其他评估中也表现出有竞争力的结果。

我们的基础模型在大多数基准测试中都表现出显著优势，我们乐观地认为，后训练技术的进步将把下一版本的 Qwen2.5-Max 提升到新的高度。

数据和模型规模的扩展不仅展示了模型智能的进步，也反映了我们对开拓性研究的坚定承诺。我们致力于通过规模化强化学习的创新应用来增强大型语言模型的思维和推理能力。

**- Alibaba Qwen 2.5 Press Release, 1/25**

## Baidu launches new AI model amid mounting competition

By Reuters

April 24, 2025 9:05 PM PDT · Updated 12 days ago



People walk near a Baidu logo at the company headquarters in Beijing, China April 23, 2021. REUTERS/Florence Lo/File Photo [Purchase Licensing Rights](#)

BEIJING, April 25 (Reuters) - Chinese search engine giant Baidu (9888.HK) launched its latest artificial intelligence model, Ernie 4.5 Turbo, on Friday, as it ramps up AI product offerings amid mounting competition in the Chinese AI market.

It also launched its new reasoning model Ernie X1 Turbo.

*ERNIE 4.5 Turbo is the newest member of the flagship ERNIE foundation model family. Imagine an AI that's not just smart, but also affordable and versatile. Here's why it's turning heads:*

*- Multimodal Prowess: It excels in handling text, images, and even videos, making it a Swiss Army knife for developers.*

*- Cost-Effectiveness: Priced at just RMB 0.8 per million tokens for input and RMB 3.2 for output, it's 80% cheaper than its predecessor – and a fraction of the cost of leading competitors. It costs only 40% of DeepSeek V3 and just 0.2% of GPT-4.5.*

*- High Performance: Benchmark tests show it matches GPT-4.1 and outperforms GPT-4o in most multimodal tasks – delivering high-impact results with every run.*

**- Baidu Post on X, 4/24/25**

Source: Reuters, 'Baidu launches new AI model amid mounting competition' (4/24/25); Baidu via X, 'Supercharging AI Innovation with More Powerful and More Affordable New Models' (4/24/25)

## Baidu launches new AI model amid mounting competition

By Reuters

April 24, 2025 9:05 PM PDT · Updated 12 days ago



People walk near a Baidu logo at the company headquarters in Beijing, China April 23, 2021. REUTERS/Florence Lo/File Photo [Purchase Licensing Rights](#)

BEIJING, April 25 (Reuters) - Chinese search engine giant Baidu (9888.HK) launched its latest artificial intelligence model, Ernie 4.5 Turbo, on Friday, as it ramps up AI product offerings amid mounting competition in the Chinese AI market.

It also launched its new reasoning model Ernie X1 Turbo.

ERNIE 4.5 Turbo 是旗舰 ERNIE 基础模型系列中的最新成员。想象一下，一个不仅聪明，而且价格实惠且用途广泛的 AI。这就是它引人注目的原因：

- 多模态能力：它擅长处理文本、图像，甚至视频，使其成为开发人员的瑞士军刀。

- 成本效益：输入价格仅为每百万 tokens 人民币 0.8 元，输出价格为人民币 3.2 元，比其前身便宜 80%–，仅为主要竞争对手成本的一小部分。它的成本仅为 DeepSeek V3 的 40%，仅为 GPT-4.5 的 0.2%。

- 高性能：基准测试表明，它在大多数多模态任务中与 GPT-4.1 相匹配，并且优于 GPT-4o – 每次运行都能提供高影响力的结果。

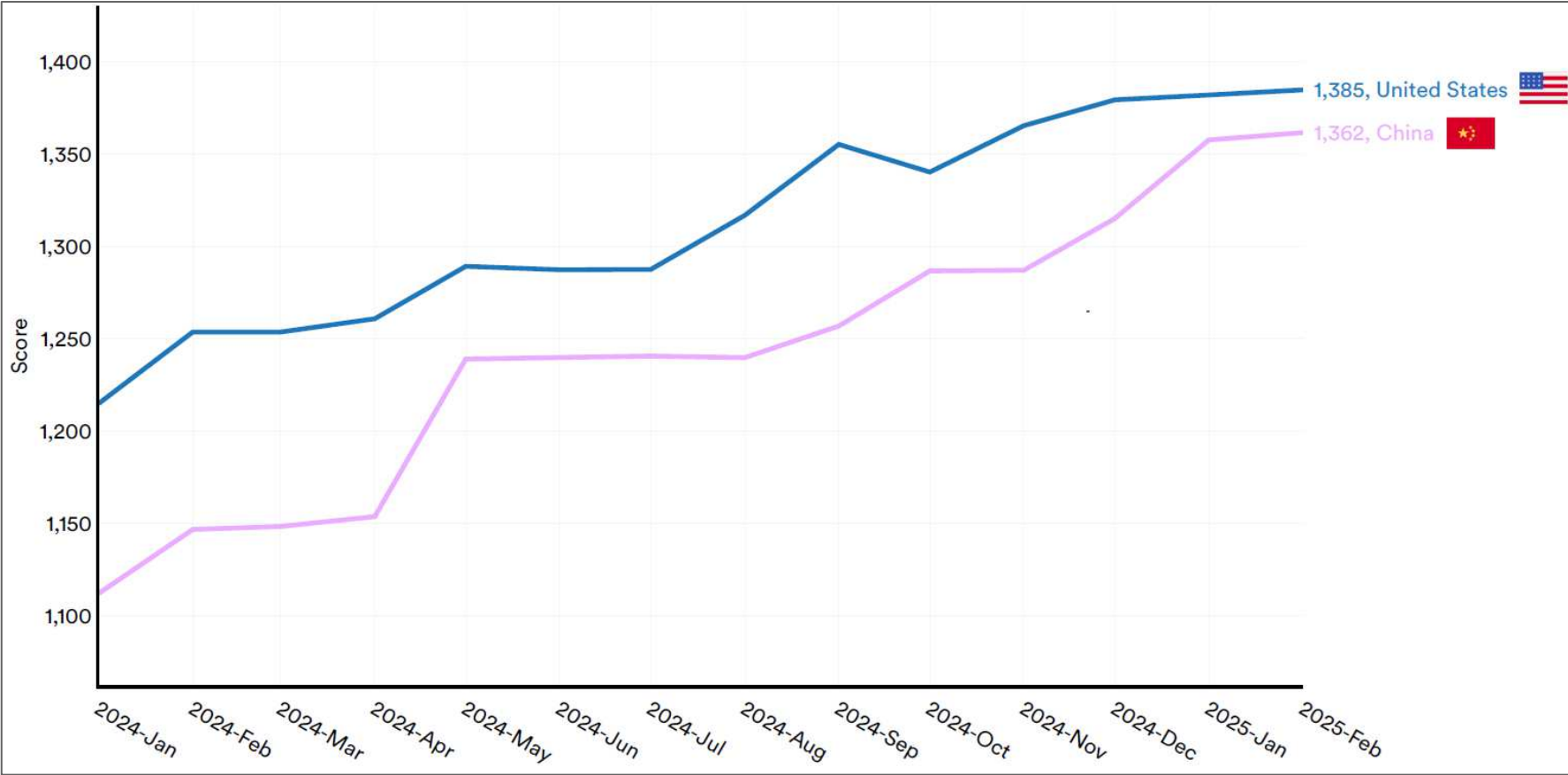
- 百度在 X 上的帖子，2025 年 4 月 24 日

Source: Reuters, 'Baidu launches new AI model amid mounting competition' (4/24/25); Baidu via X, 'Supercharging AI Innovation with More Powerful and More Affordable New Models' (4/24/25)



# China AI = LLM Performance Catching Up to USA Models, per Stanford HAI...

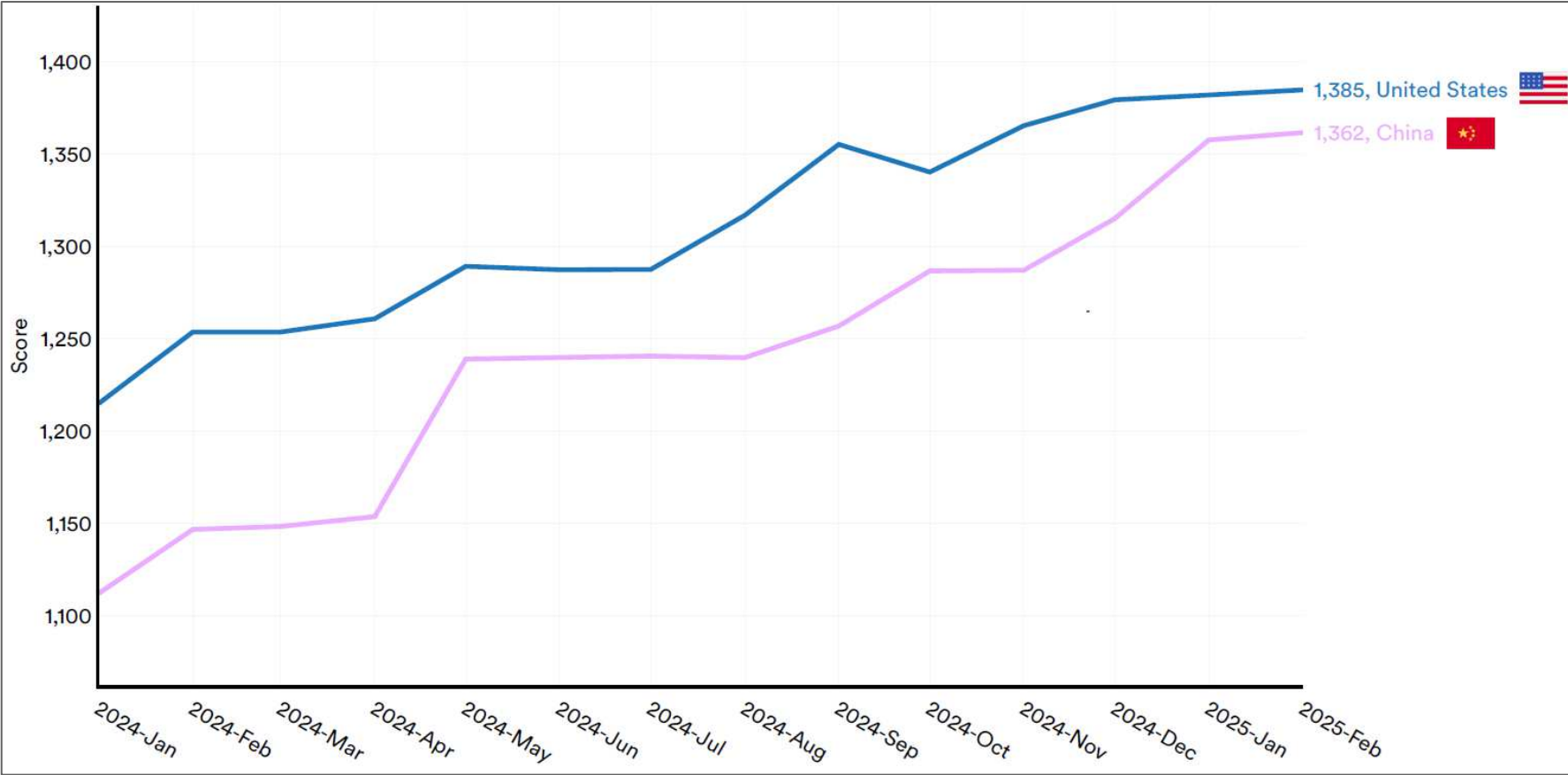
Performance of Top-Scoring USA vs. Chinese AI Model on LMSYS Chatbot Arena – 1/24-2/25, per Stanford HAI & LMSYS



Note: The LMSYS Chatbot Arena is a public website where people compare two AI chatbots by asking them the same question and voting on which answer is better. The results help rank how well different language models perform based on human judgment. Only the highest-scoring model in any given month is shown in this comparison. Source: LMSYS via Nestor Maslej et al., 'The AI Index 2025 Annual Report,' AI Index Steering Committee, Stanford HAI (4/25)

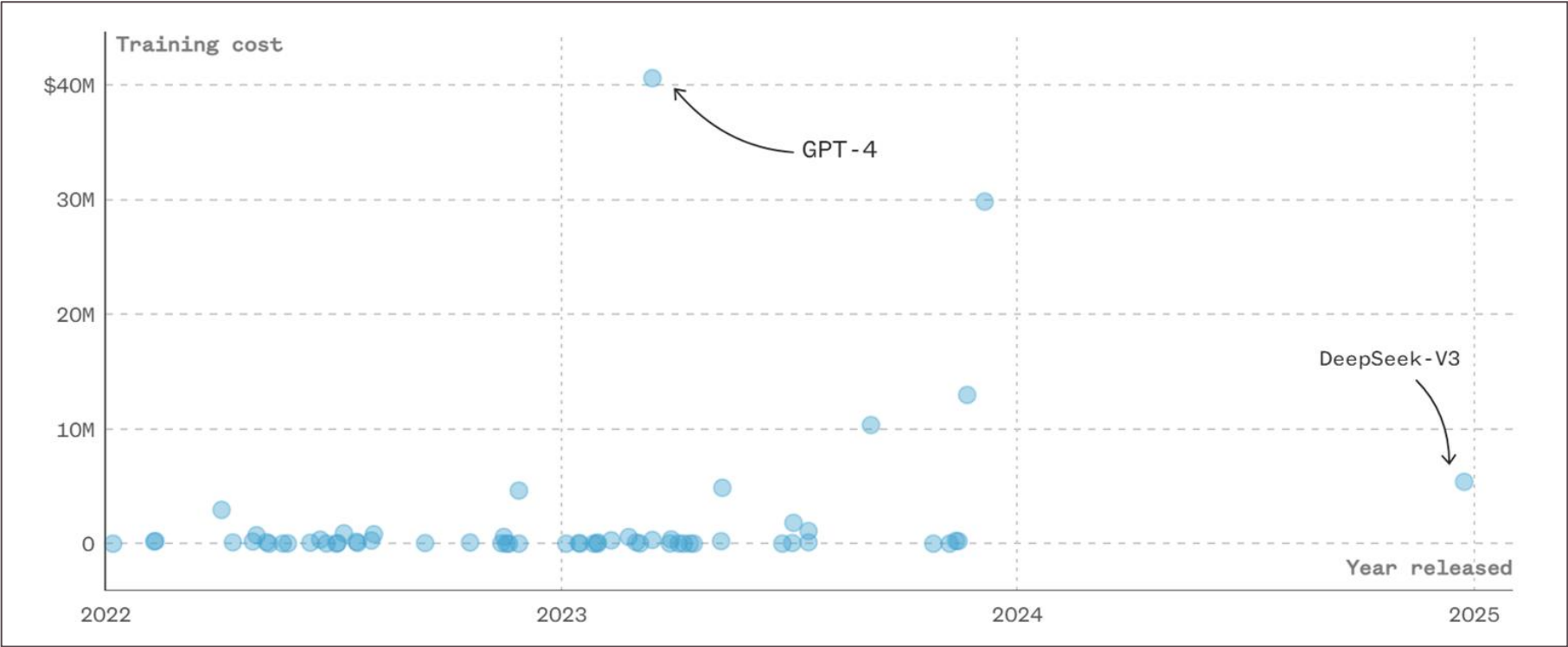
# China AI =LLM 性能追赶美国模型，根据斯坦福 HAI...

LMSYS Chatbot Arena 上得分最高的美国与中国 AI 模型的性能比较 – 1/24-2/25，根据斯坦福 HAI 和 LMSYS

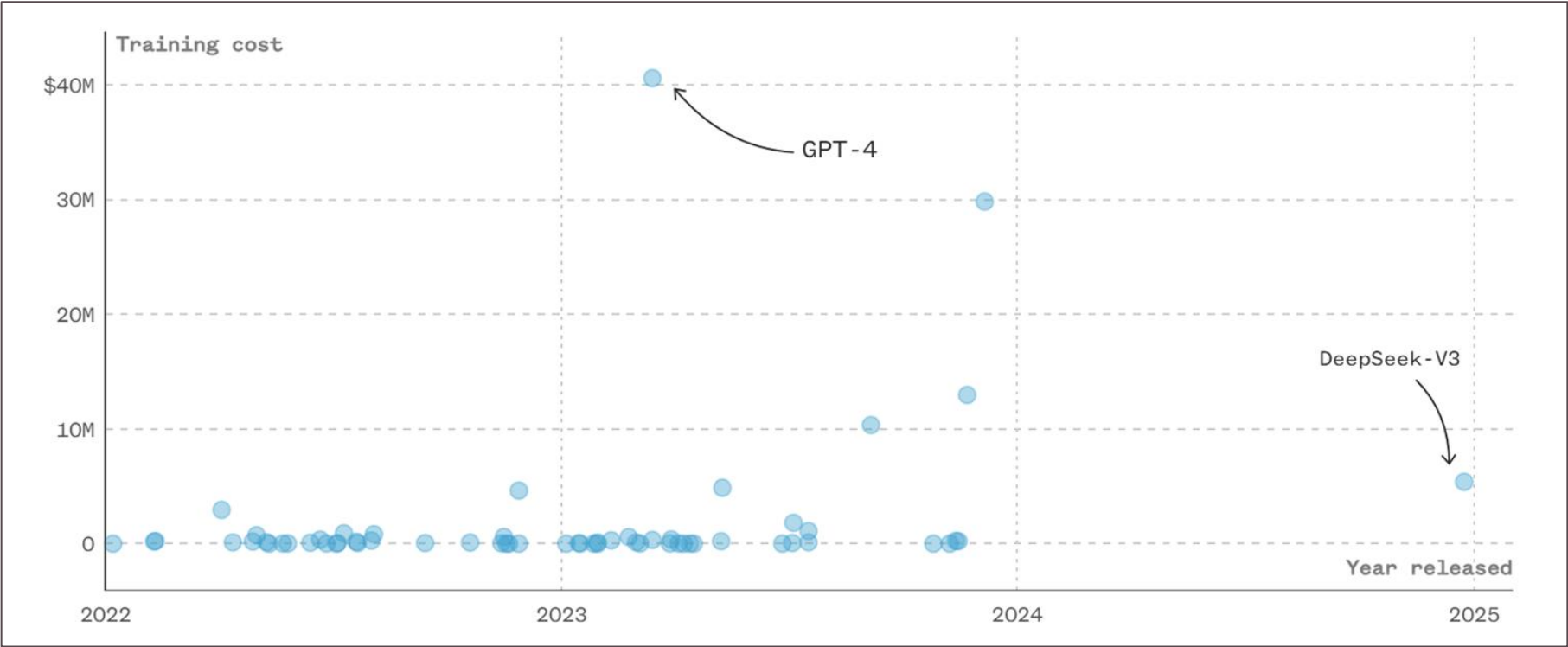


注意：LMSYS Chatbot Arena 是一个公共网站，人们通过向两个 AI 聊天机器人提出相同的问题并投票选出哪个答案更好来比较它们。结果有助于根据人类的判断来评估不同语言模型的表现。在此比较中，仅显示任何给定月份中得分最高的模型。来源：LMSYS 通过 Nestor Maslej et al., 'The AI Index 2025 Annual Report,' AI Index Steering Committee, Stanford HAI (4/25)

LLM Training Cost by Year Released – 2022-2024, per Epoch AI & NBC News



LLM Training Cost by Year Released – 2022-2024, per Epoch AI & NBC News



## Huawei delivers advanced AI chip ‘cluster’ to Chinese clients cut off from Nvidia

CloudMatrix 384 sent to data centres serving Chinese tech companies as they seek to fill gap created by US export controls



Huawei is accelerating its chip development to help Chinese tech companies compete with global peers © Tobias Schwarz/AFP/Getty Images

Zijing Wu in Hong Kong and Eleanor Olcott in Beijing

Published APR 29 2025



*Huawei has started the delivery of its advanced artificial intelligence chip ‘cluster’ to Chinese clients who are increasing orders after being cut off from Nvidia’s semiconductors because of Washington’s export restrictions...*

**- Financial Times, 4/29/25**

## Huawei delivers advanced AI chip ‘cluster’ to Chinese clients cut off from Nvidia

CloudMatrix 384 sent to data centres serving Chinese tech companies as they seek to fill gap created by US export controls



Huawei is accelerating its chip development to help Chinese tech companies compete with global peers © Tobias Schwarz/AFP/Getty Images

Zijing Wu in Hong Kong and Eleanor Olcott in Beijing

Published APR 29 2025



华为已开始交付其先进的人工智能芯片“集群”给因华盛顿的出口限制而与英伟达半导体断绝关系后，订单不断增加的中国客户 ...

- 金融时报，25 年 4 月 29 日

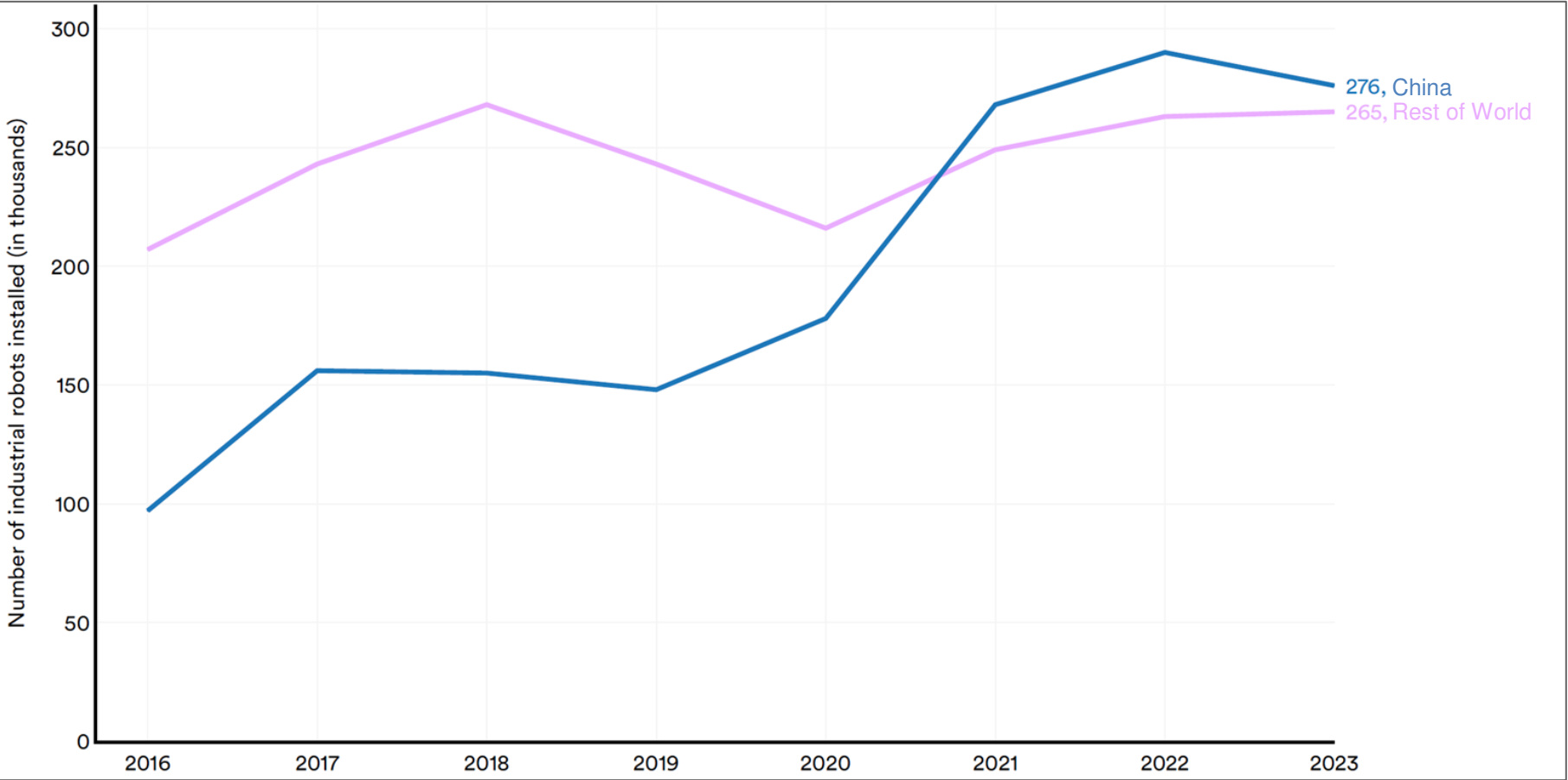
Source: Financial Times, ‘Huawei delivers advanced AI chip ‘cluster’ to Chinese clients cut off from Nvidia’ (4/29/25)

来源：金融时报，“华为向与英伟达断绝关系的中国客户交付先进的 AI 芯片‘集群’”(25 年 4 月 29 日)

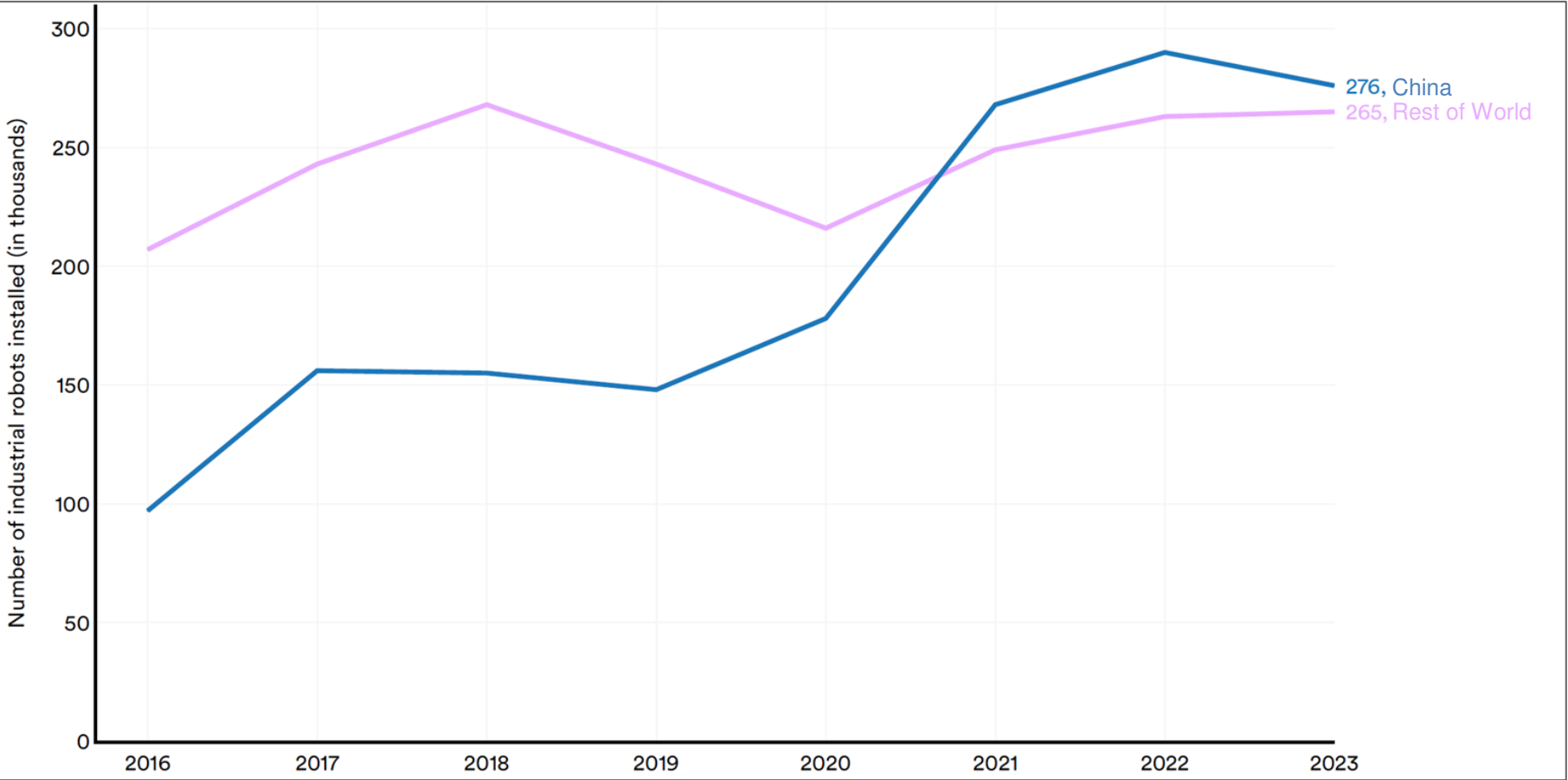
...China AI = Industrial Robot Installed Base Higher vs. Rest of World...

… 中国人工智能 = 工业机器人安装基数高于世界其他地区 …

Number of Industrial Robots Installed (China vs. Rest of World) (K) – 2023, per IFR



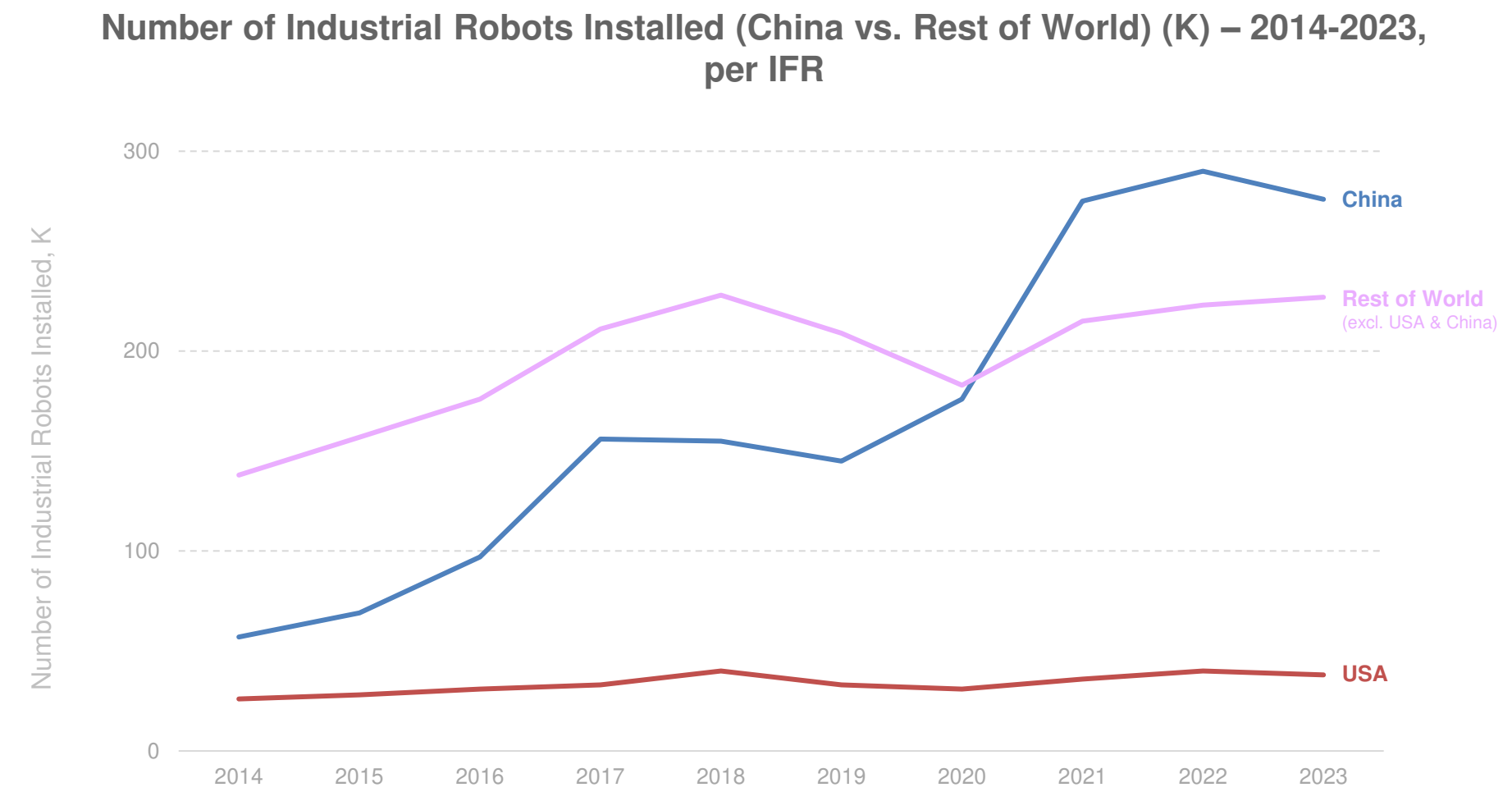
Number of Industrial Robots Installed (China vs. Rest of World) (K) – 2023, per IFR



Source: International Federation of Robotics (IFR) (2024) via Nestor Maslej et al., 'The AI Index 2025 Annual Report,' AI Index Steering Committee, Stanford HAI (4/25)

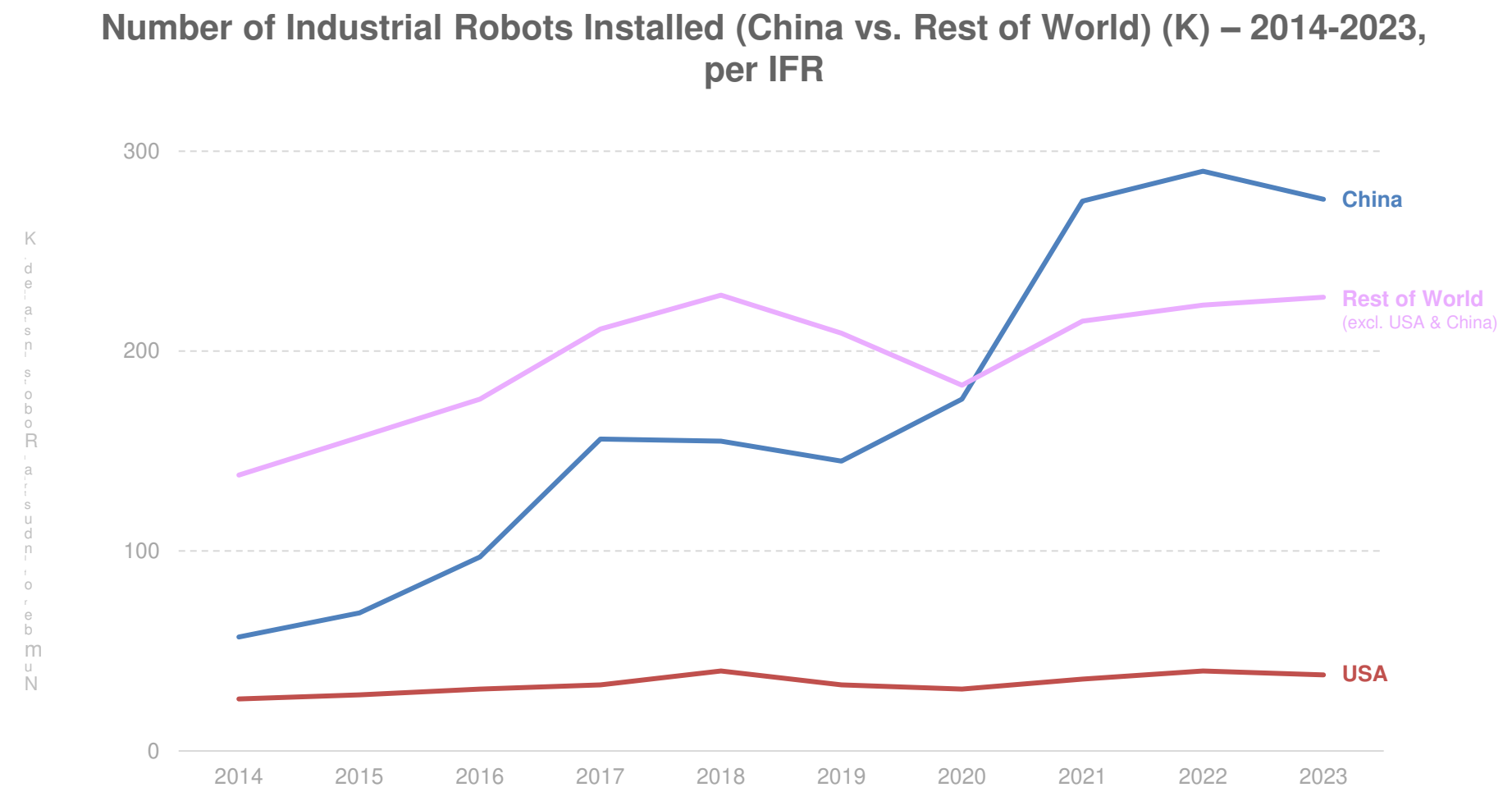
Source: International Federation of Robotics (IFR) (2024) via Nestor Maslej et al., 'The AI Index 2025 Annual Report,' AI Index Steering Committee, Stanford HAI (4/25)

...China AI =  
Industrial Robot Installed Base Higher vs. Rest of World



Source: International Federation of Robotics (IFR) (2024)

...China AI =  
Industrial Robot Installed Base Higher vs. Rest of World



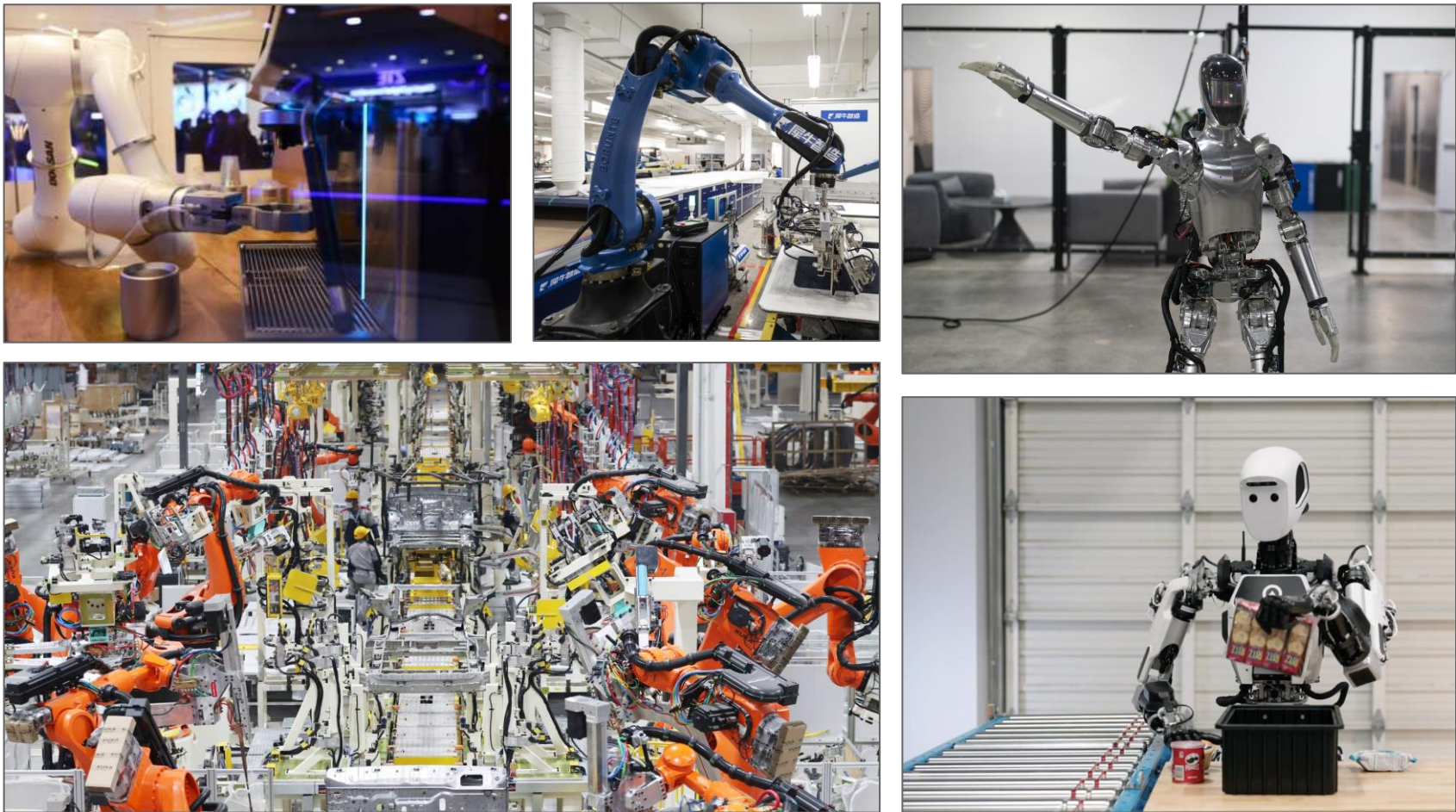
Source: International Federation of Robotics (IFR) (2024)



Robots – Industrial & Humanoid =  
Creating New Data @ New Scale

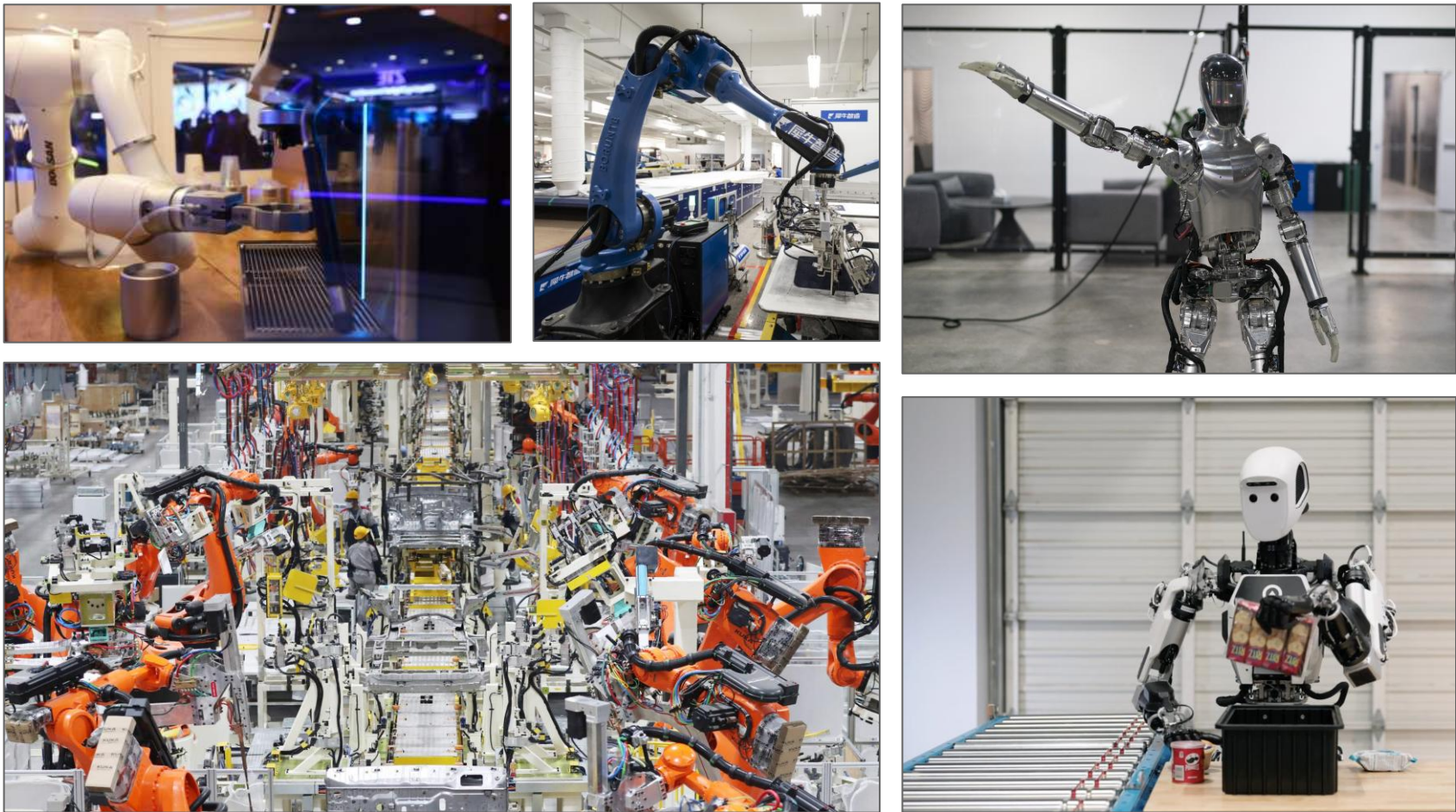
机器人 – 工业和人形机器人 = 以新的  
规模创建新数据

Images of Industrial & Humanoid Robots, per *The Wall Street Journal*



Source: The Wall Street Journal (2/18, 5/22, 9/22, 5/25)

Images of Industrial & Humanoid Robots, per *The Wall Street Journal*



Source: The Wall Street Journal (2/18, 5/22, 9/22, 5/25)

China Consumer AI Usage =

DeepSeek Rose Quickly

中国消费者人工智能使用情况 =

DeepSeek Rose Quickly



To understand how the generative AI market is evolving, it helps to examine the divergence in provider usage across regions, channels, and user preferences. At a global level, OpenAI’s ChatGPT remains the clear leader in both desktop and mobile user share. But underneath the surface, the market is shifting. Platforms like Anthropic’s Claude are gaining momentum, and Google’s Gemini continues to grow. xAI’s Grok posted a staggering +294% increase in global website visits month-over-month according to Similarweb – making it the fastest-growing AI assistant during the 2/25-3/25 window.

Geography is also playing an increasingly central role in shaping which models win. ChatGPT dominates in most countries – excluding Russia and China, where ChatGPT cannot operate and DeepSeek is strong.

China users are turning to local models at scale. According to Roland Berger Consulting, the top 10 AI apps by monthly active users in China are domestically developed...DeepSeek, Kimi, Nami AI, and ERNIE Bot are each racking up tens of millions of users. The story is different outside China, where ChatGPT leads by a wide margin. The bifurcation is clear: domestic champions are emerging in China, while global platforms dominate elsewhere. This reflects differences in regulation, language, cultural alignment, and platform reach.

It’s foundational to remember how China has restricted platform access in its country. Facebook, Twitter, Google and YouTube have been unavailable to Chinese citizens since 2010 or earlier. Other restricted platforms include the likes of Instagram, WhatsApp, Wikipedia, Telegram and Spotify, and more recently, the likes of ChatGPT, Google Gemini, Anthropic Claude, Meta AI and Microsoft Copilot.

Sentiment is varied too. According to Stanford HAI and Ipsos, China citizens are materially more optimistic about AI’s net benefits than their USA counterparts. 83% of Chinese respondents in 2024 said AI products and services have more benefits than drawbacks – up from 78% in 2022. In contrast, only 39% of USA respondents shared that view, with little change over the two-year period.

It also reflects a deeper philosophical divide in how societies are adapting to AI: not just who builds it, but how it’s perceived and embraced. In this environment, platform choice isn’t just about price or performance. It may be increasingly shaped by national identity.

为了解生成式人工智能市场的演变情况，有必要考察不同地区、渠道和用户偏好中提供商使用情况的差异。在全球范围内，OpenAI 的 ChatGPT 仍然是桌面和移动用户份额方面当之无愧的领导者。但在表面之下，市场正在发生变化。Anthropic 的 Claude 等平台正在获得发展势头，而谷歌的 Gemini 则继续增长。根据 Similarweb 的数据，xAI 的 Grok 的全球网站访问量环比增长了惊人的 +294%——使其成为 2/25-3/25 窗口期增长最快的人工智能助手。

地域在塑造哪些模型胜出方面也发挥着越来越重要的作用。ChatGPT 在大多数国家 / 地区占据主导地位 —— 俄罗斯和中国除外，ChatGPT 无法在这些国家 / 地区运营，而 DeepSeek 则表现强劲。

中国用户正在大规模转向本地模型。根据罗兰贝格咨询公司的数据，中国月度活跃用户最多的 10 款人工智能应用均为国内开发 ……DeepSeek、Kimi、Nami AI 和 ERNIE Bot 每款应用都在积累数千万用户。中国以外的情况则不同，ChatGPT 以巨大优势领先。这种分化非常明显：国内巨头正在中国崛起，而全球平台则在其他地区占据主导地位。这反映了法规、语言、文化契合度和平台影响力的差异。

重要的是要记住中国如何限制其国内的平台访问。自 2010 年或更早以来，中国的公民就无法使用 Facebook、Twitter、Google 和 YouTube。其他受限制的平台包括 Instagram、WhatsApp、Wikipedia、Telegram 和 Spotify，以及最近的 ChatGPT、Google Gemini、Anthropic Claude、Meta AI 和 Microsoft Copilot。

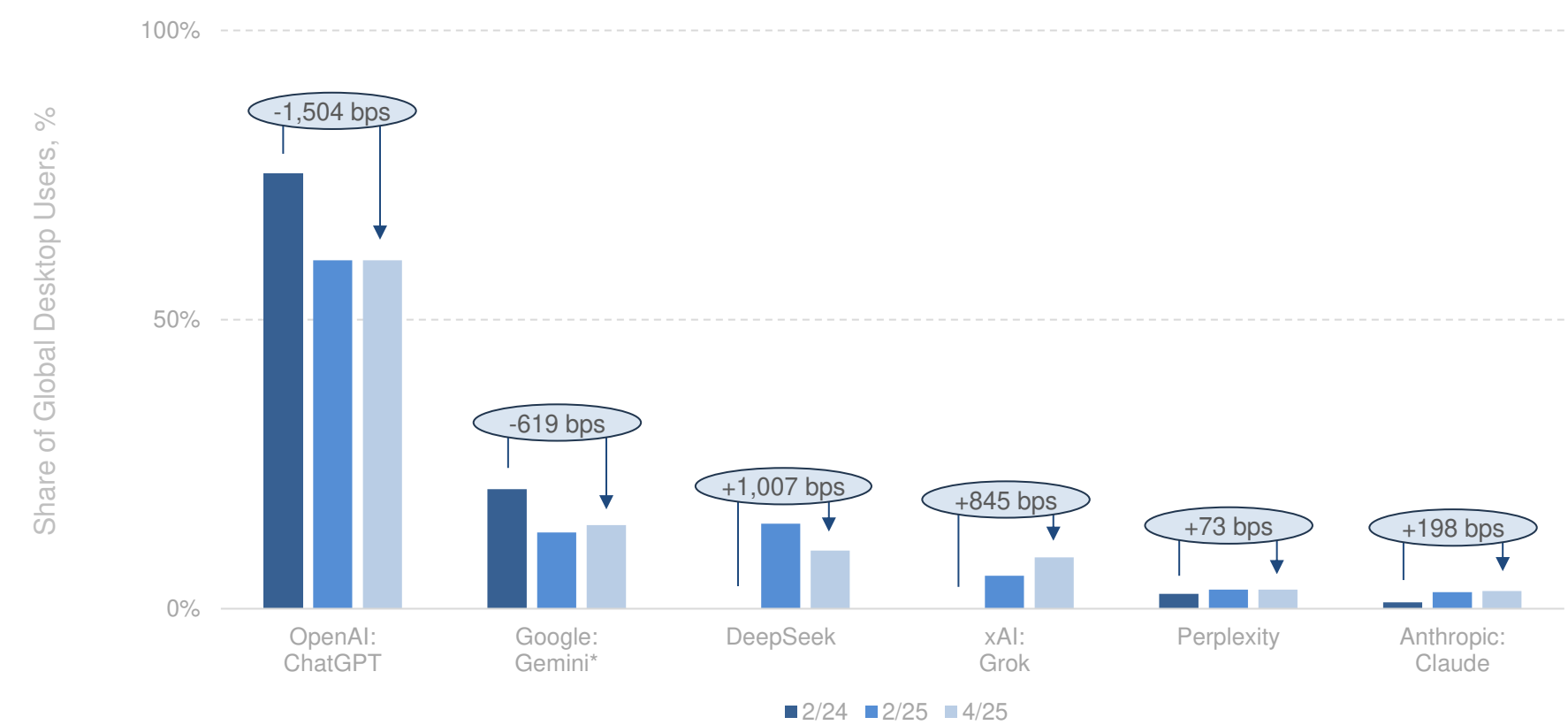
情绪也各不相同。根据斯坦福 HAI 和益普索的说法，中国公民对人工智能的净收益的乐观程度明显高于他们的美国同行。2024 年，83% 的中国受访者表示，人工智能产品和服务的益处大于弊端 —— 高于 2022 年的 78%。相比之下，只有 39% 的美国受访者持有同样的观点，两年期间几乎没有变化。

这也反映了社会在适应人工智能方面更深层次的哲学分歧：不仅是谁构建了它，而且是如何看待和接受它的。在这种环境下，平台选择不仅仅是价格或性能的问题。它可能会越来越受到民族认同的影响。

# LLM User Share – Desktop Users = OpenAI ChatGPT Leads...DeepSeek Rose Quickly, per YipitData...

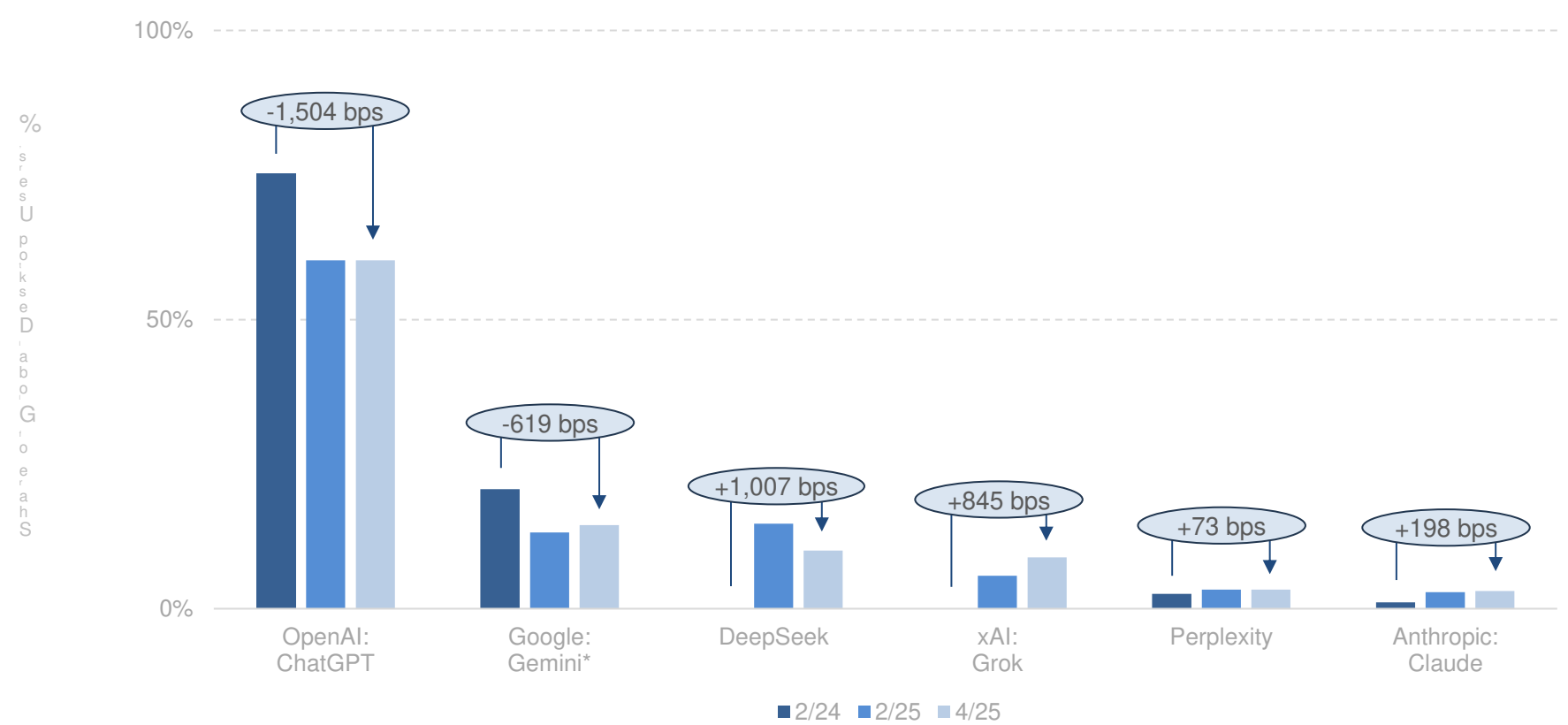
# LLM User Share – Desktop Users =OpenAI ChatGPT 领先 …… 根据 YipitData……， DeepSeek 迅速崛起

Estimated Global Monthly Active Desktop User Share – 2/24-4/25, per YipitData



*\*Chatbot only. Does not include other places Gemini is integrated. Note: User share shown across these five providers; other LLMs' user share not shown. Desktop users only. Figures calculate the number of users on a given platform, divided by the number of users on all platforms combined. Figures are non-deduped (i.e., users using multiple platforms may be counted twice). Data is a subset of global internet users and absolute user data will be understated; however, given that the panel is globally-representative (with limitations on China-specific data), relative comparisons / trends are informative. Data measures several million global active desktop users' clickstream data. Data consists of users' web requests & is collected from web services / applications, such as VPNs and browser extensions. Panel is globally-representative (with limitations on China-specific data). Users must have been part of the panel for 2 consecutive months to be included. Data is non-deduped; i.e., some users may use multiple platforms. Source: YipitData (accessed 5/25)*

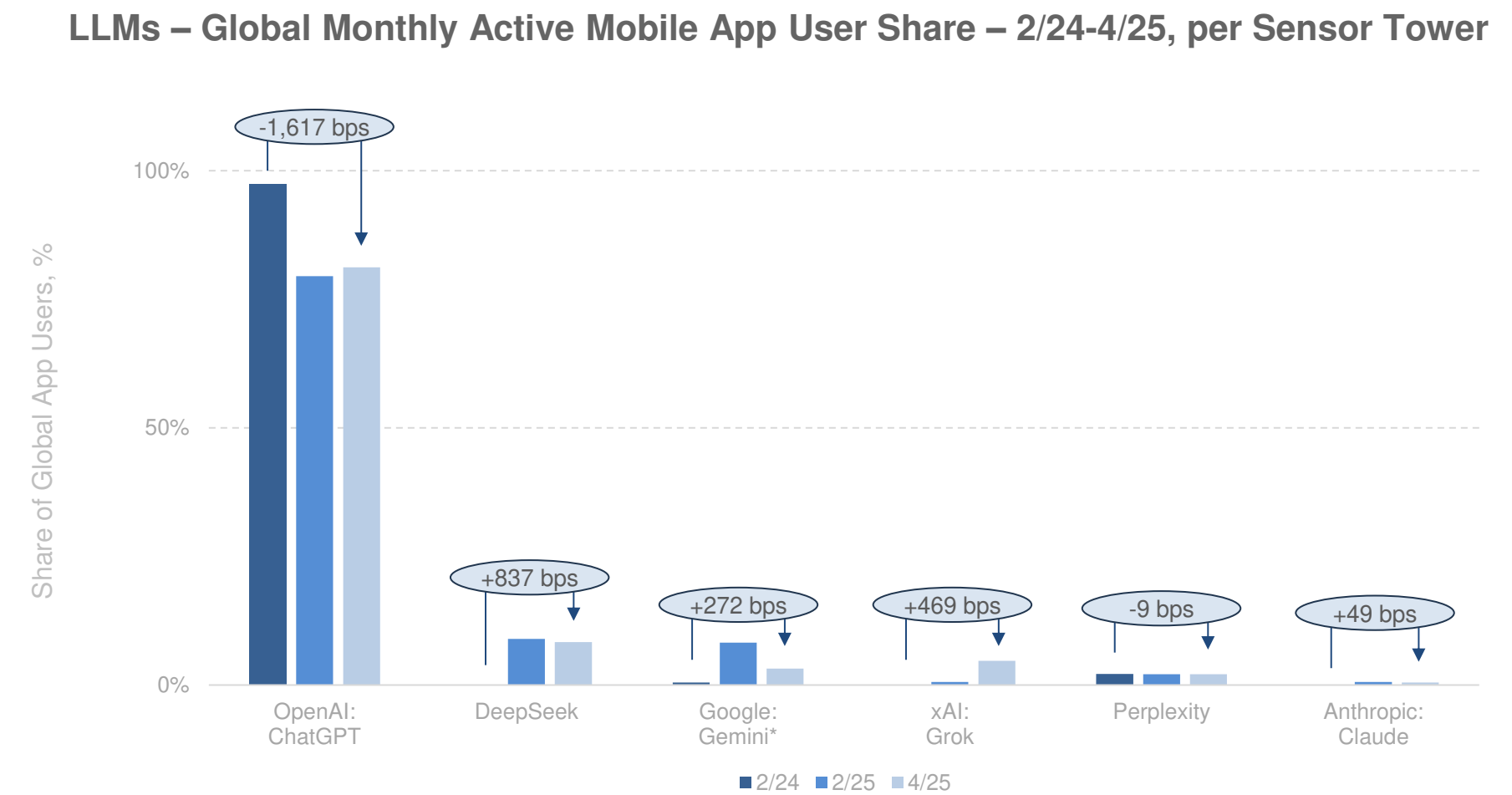
Estimated Global Monthly Active Desktop User Share – 2/24-4/25, per YipitData



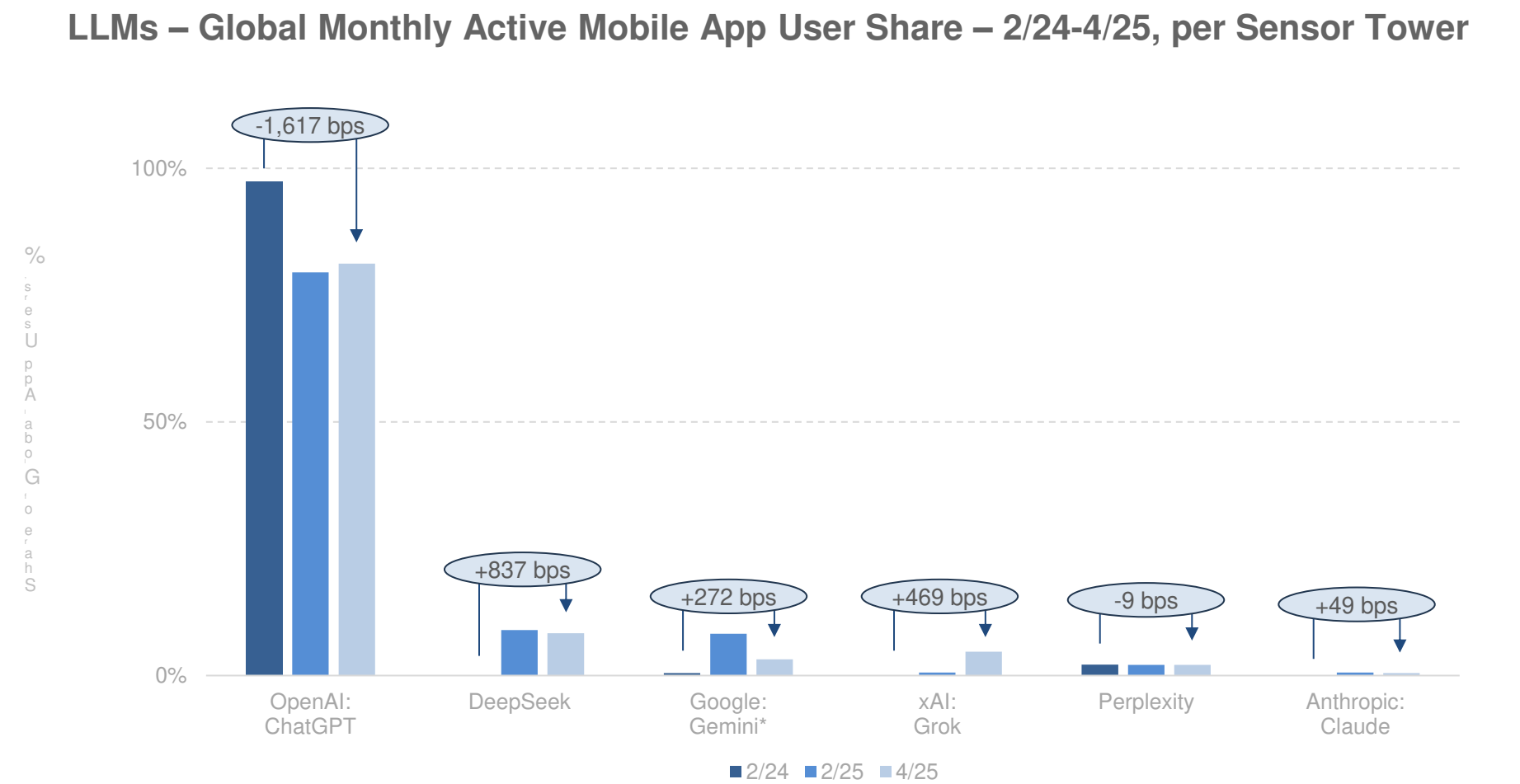
*\* 仅限聊天机器人。不包括 Gemini 集成的其他位置。注意：用户份额显示在这五个提供商之间；其他 LLM 的用户份额未显示。仅限桌面用户。数据计算给定平台上的用户数量，除以所有平台上用户数量的总和。数据未经去重（即，使用多个平台的用户可能会被计算两次）。数据是全球互联网用户的一个子集，绝对用户数据将被低估；但是，鉴于该小组具有全球代表性（对中国特定数据存在限制），相对比较 / 趋势具有参考价值。数据衡量了数百万全球活跃桌面用户的点击流数据。数据包括用户的网络请求，并且是从网络服务 / 应用程序（例如 VPN 和浏览器扩展）收集的。该小组具有全球代表性（对中国特定数据存在限制）。用户必须连续 2 个月成为该小组的成员才能被纳入。数据未经去重；即，某些用户可能使用多个平台。来源：YipitData（访问日期 5/25）*

...LLM User Share – Mobile App Users = OpenAI ChatGPT Leads...DeepSeek Rose Quickly, per Sensor Tower...

…LLM 用户份额 – 移动应用用户=OpenAI ChatGPT 领先 … 根据 Sensor Tower，DeepSeek 迅速崛起 …



*\*Chatbot only. Does not include other places Gemini is integrated. Note: User share shown across these five providers; other LLMs' user share not shown. China data may be incomplete due to reporting gaps. ChatGPT app not available in China, Russia and select other countries as of 5/25. Data is non-deduped; i.e., some users may use multiple platforms. Data for standalone apps only. Source: Sensor Tower (accessed 5/25)*



*\* 仅限聊天机器人。不包括 Gemini 集成的其他位置。注意：用户份额显示在这五个提供商之间；其他 LLM 的用户份额未显示。由于报告缺失，中国的数据可能不完整。截至 5/25，ChatGPT 应用程序在中国、俄罗斯和某些其他国家 / 地区不可用。数据未经重复数据删除；即，某些用户可能使用多个平台。仅限独立应用程序的数据。来源：Sensor Tower（访问时间：5/25）*

...LLM User Share – Mobile App Downloads + Users = ChatGPT Supporting Strong Momentum...

Global Statistics on Apple App Store + Google Play Store – 2/25-4/25, per Sensor Tower

		Downloads (MM)			MAUs (MM)		
		2/25	3/25	4/25	2/25	3/25	4/25
LLM Apps	ChatGPT	56MM	80MM	124MM	378MM	432MM	530MM
	DeepSeek	34	20	18	43	48	55
	Grok	4	14	16	3	16	31
	Gemini*	16	17	15	20	21	21
	Perplexity	3	4	4	10	12	14
	Claude	1	1	1	3	4	3
‘Traditional’ Apps	YouTube	13	10	9	2,799	2,805	2,809
	Google Chrome	9	9	7	2,369	2,380	2,387
	Facebook	46	47	45	2,104	2,110	2,103

*\*Chatbot only. Does not include other places Gemini is integrated. Note: China data may be incomplete due to reporting gaps. ChatGPT app not available in China, Russia and select other countries as of 5/25. Data is non-deduped; i.e., some users may use multiple platforms. Data for standalone apps only. Source: Sensor Tower (accessed 5/25)*

…LLM 用户份额 – 移动应用下载量 + 用户 =ChatGPT 强劲势头不减 …

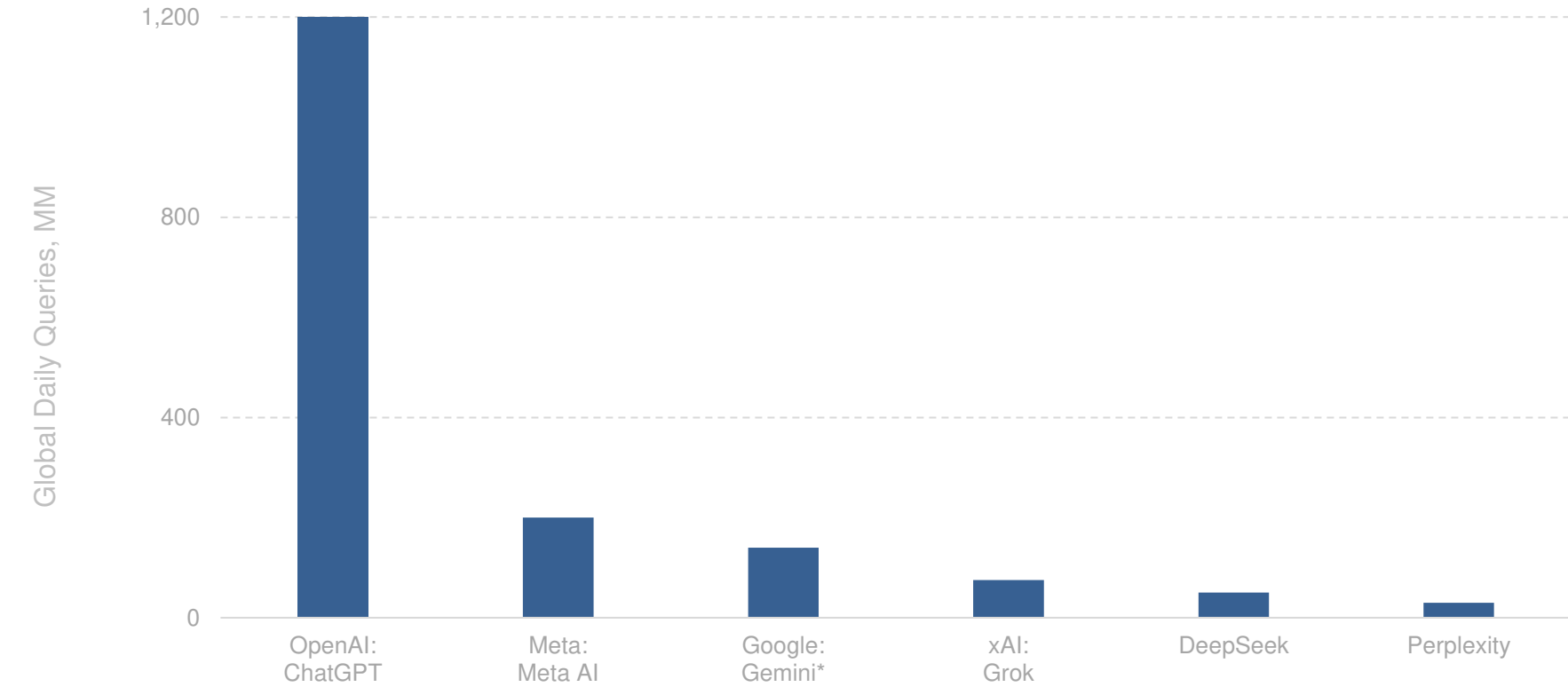
Global Statistics on Apple App Store + Google Play Store – 2/25-4/25, per Sensor Tower

		下载量 ( 百万 )			MAUs (MM)		
		2/25	3/25	4/25	2/25	3/25	4/25
LLM Apps	ChatGPT	56MM	80MM	1.24 亿	3.78 亿	432MM	530MM
	DeepSeek	34	20	18	43	48	55
	Grok	4	14	16	3	16	31
	Gemini*	16	17	15	20	21	21
	Perplexity	3	4	4	10	12	14
	Claude	1	1	1	3	4	3
“ 传统 ” Apps	YouTube	13	10	9	2,799	2,805	2,809
	Google Chrome	9	9	7	2,369	2,380	2,387
	Facebook	46	47	45	2,104	2,110	2,103

*\* 仅限聊天机器人。不包括 Gemini 集成的其他位置。注意：由于报告存在差距，中国数据可能不完整。截至 5 月 25 日，ChatGPT 应用程序在中国、俄罗斯和某些其他国家 / 地区不可用。数据未去重；即，某些用户可能使用多个平台。仅限独立应用程序的数据。来源：Sensor Tower （5 月 25 日访问）*

# ...LLM User Share – Query Volume = OpenAI ChatGPT Leads, per Google

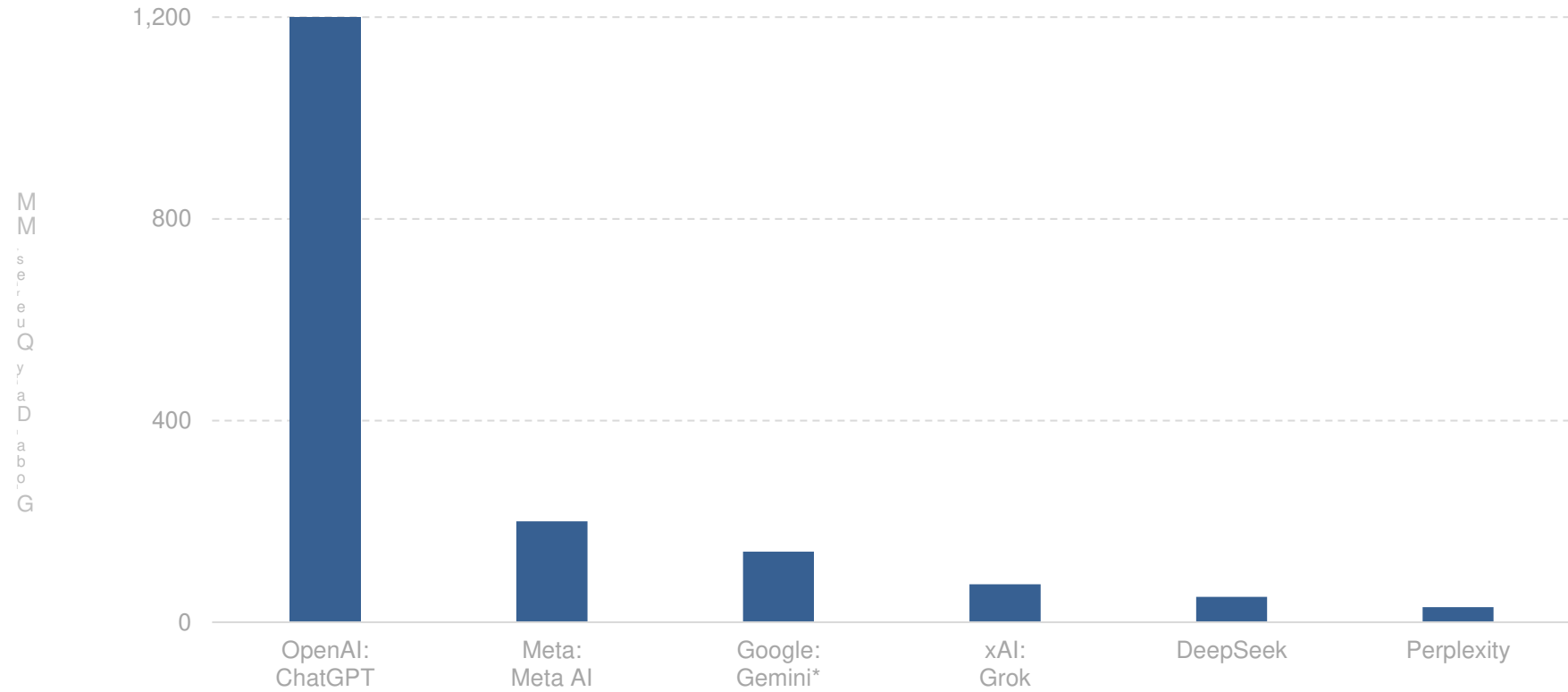
LLMs – Global Daily Query Volume (MM) – 3/28/25, per Google



*\*Chatbot only. Does not include other places Gemini is integrated. Note: DeepSeek data excludes China usage. Figures are rounded. Meta AI data quoted as '>200M.' Source: Google disclosed during testimony given in the remedies phase of 'United States v. Google LLC' (1/24/23-4/17/25). Data derived from company disclosures, Sensor Tower, AppAnnie, Similarweb, & market intelligence estimates, as reported by Business Insider, 'Google's Gemini usage is skyrocketing, but rivals like ChatGPT and Meta AI are still blowing it out of the water' (4/25) ([link](#))*

# ...LLM User Share – 查询量 = OpenAI ChatGPT 领先，根据 Google

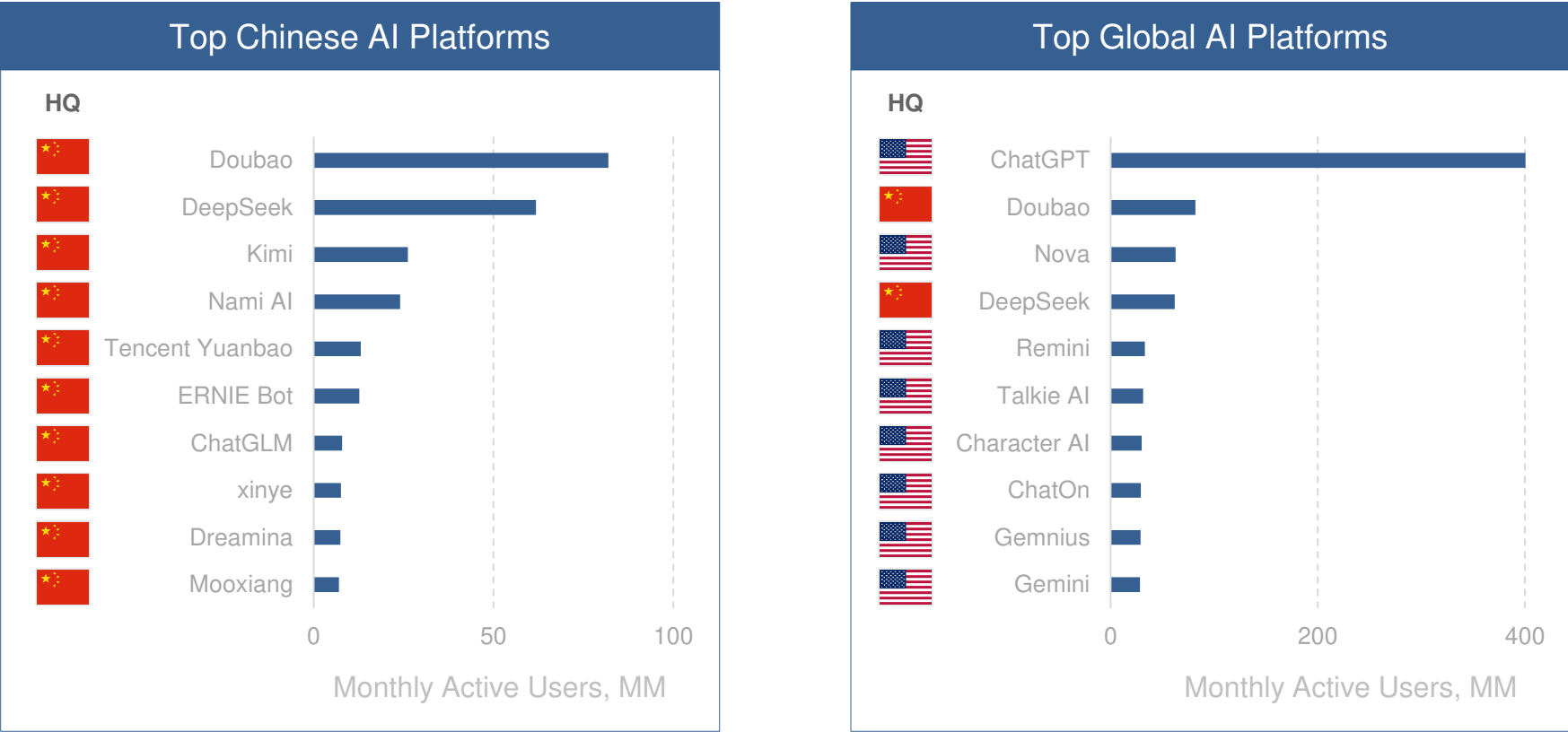
LLMs – 全球每日查询量 ( 百万 ) – 3/28/25， 根据 Google



*\* 仅限聊天机器人。不包括 Gemini 集成的其他位置。注意：DeepSeek 数据不包括中国的使用情况。数字已四舍五入。Meta AI 数据引用为 '>200M.' 来源：谷歌在 'United States v. Google LLC' (1/24/23-4/17/25) 补救阶段的证词中披露。数据来自公司披露、Sensor Tower、AppAnnie、Similarweb 和市场情报估计，如 Business Insider 报道，“谷歌的 Gemini 使用量正在飙升，但像 ChatGPT 和 Meta AI 等竞争对手仍然遥遥领先” 水 ’ (4/25) ([link](#))*

# China AI Users = Using Local AI Platforms, per Roland Berger Consulting

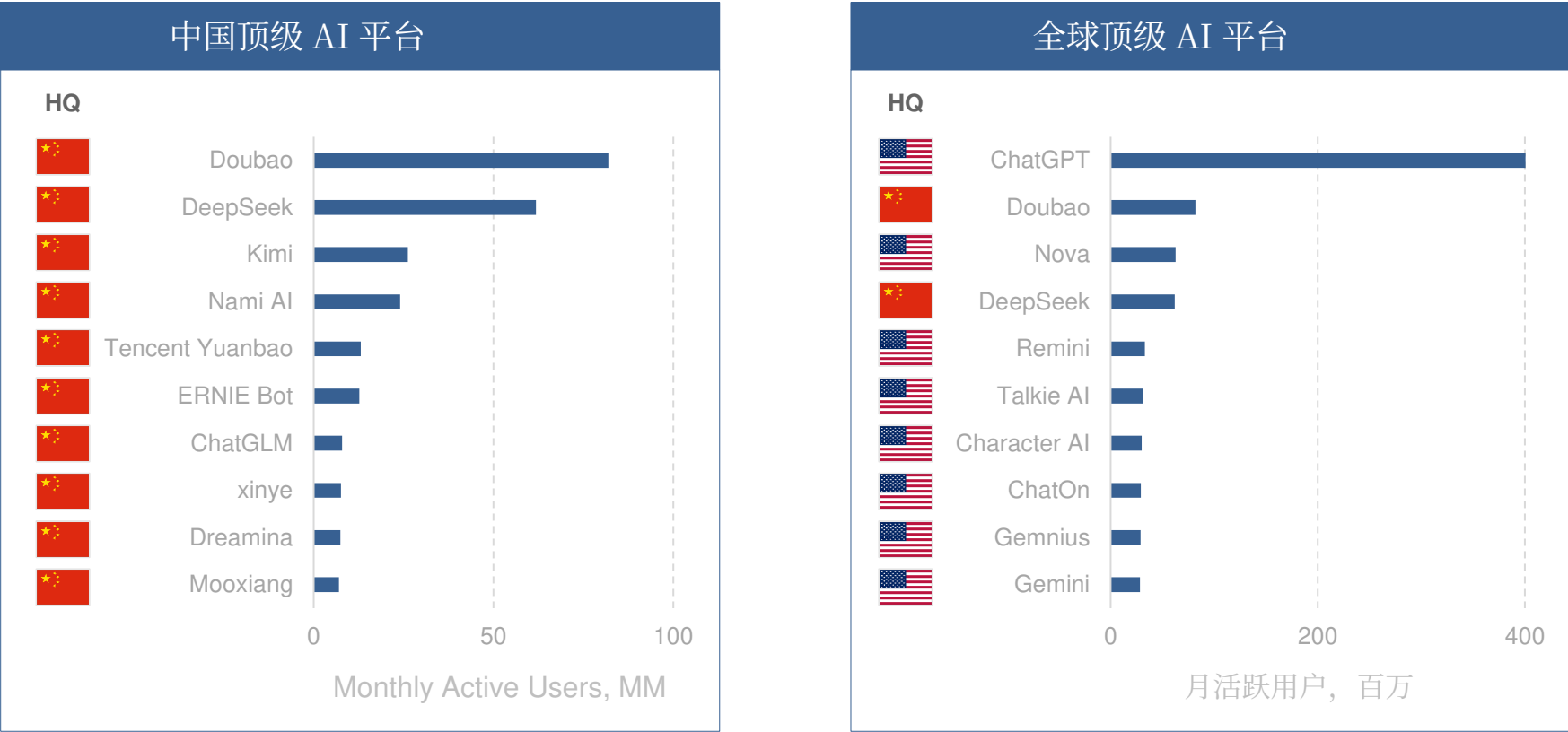
AI Platforms – Monthly Active Users (MM), China vs. Global – 3/25,  
per Roland Berger Consulting



Note: HQ = Headquarters. Axes for two charts are to different scales.  
Source: Roland Berger via AICPB, 'Five key trends in China's generative AI market in 2025' (3/25); China National Bureau of Statistics (1/25); USA Census Bureau (4/25)

# China AI Users = Using Local AI Platforms, per Roland Berger Consulting

AI 平台 – 月活跃用户（百万），中国 vs. 全球 – 3/25，根据 Roland Berger Consulting

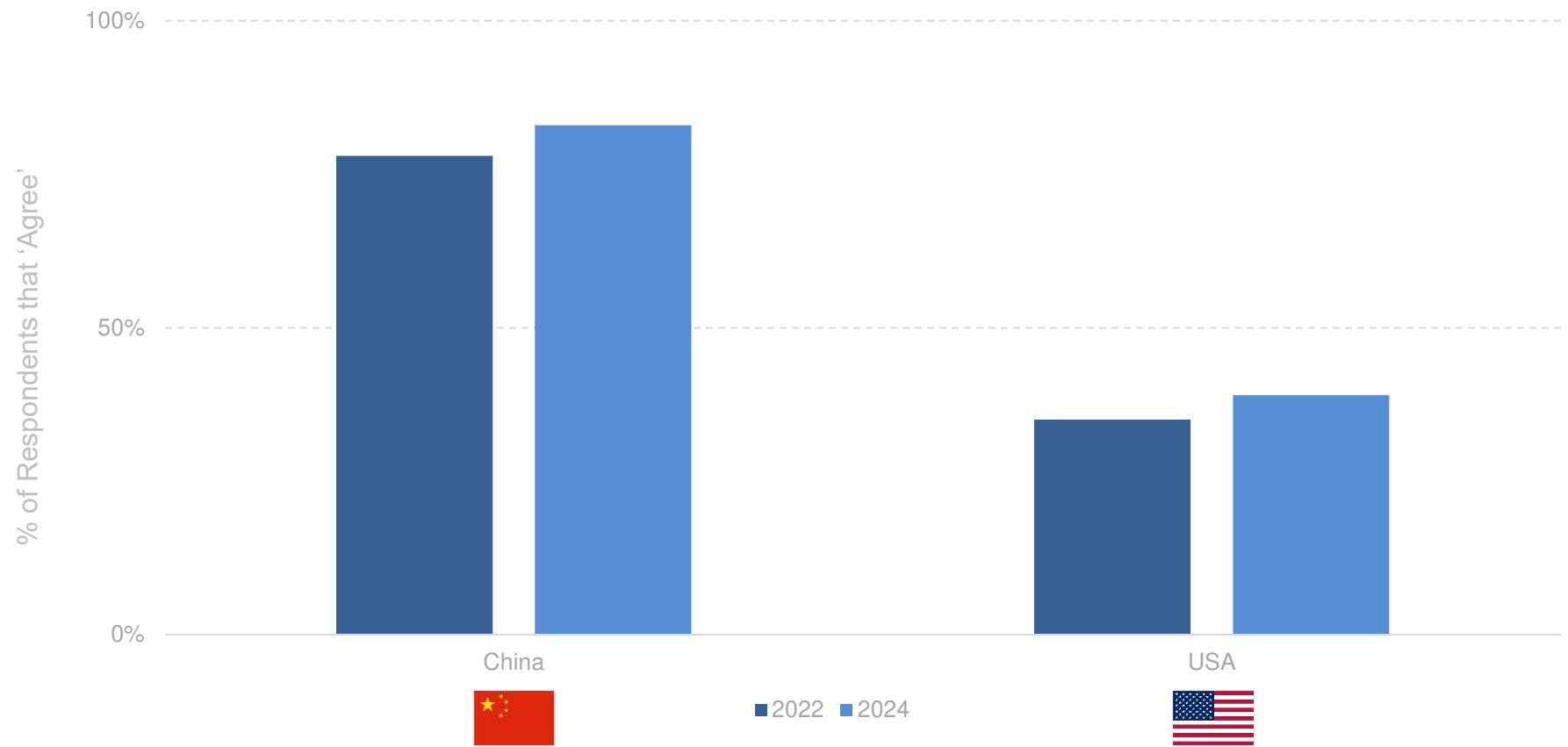


注：总部 = 。两个图表的轴采用不同的比例。来源：Roland Berger 经 AICPB，"中国生成式 AI 市场 2025 年的五大趋势" (3/25)；中国国家统计局 (1/25)；美国人口普查局 (4/25) 统计

# AI Benefits vs. Drawbacks – China vs. USA Citizens = China Materially More Optimistic Regarding Benefits

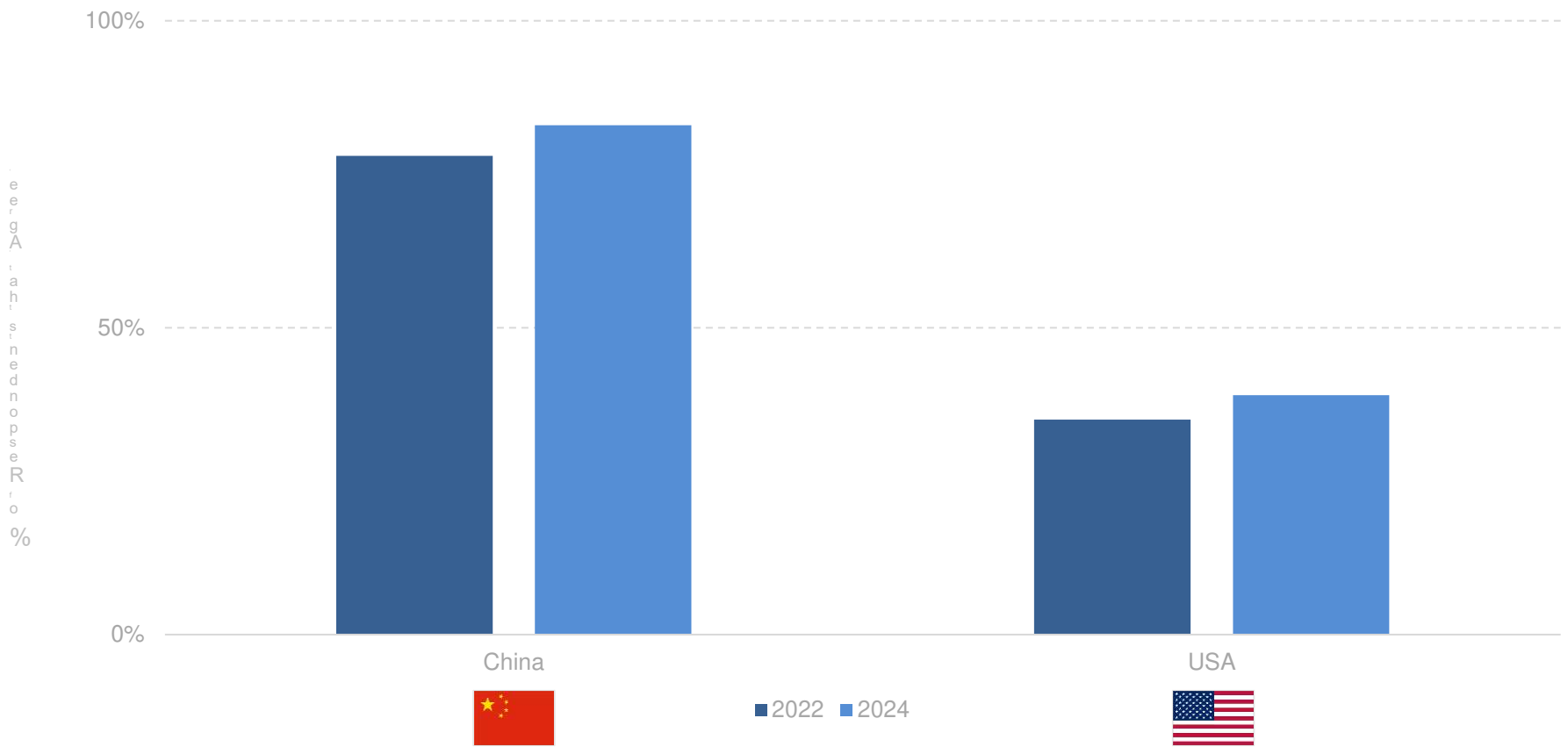
# AI 益处与弊端 – 中国与美国公民 = 中国对益处更为乐观

‘Products & Services Using AI Have More Benefits than Drawbacks’ – 2022-2024, per Stanford HAI & Ipsos



Note: N = 19,504 online adults aged 16-74 across 28 countries.  
Source: Ipsos, 'AI Monitor 2024' (6/24) as quoted in Nestor Maslej et al., 'The AI Index 2025 Annual Report,' AI Index Steering Committee, Stanford HAI (4/25)

‘Products & Services Using AI Have More Benefits than Drawbacks’ – 2022-2024, per Stanford HAI & Ipsos



注：N = 19,504 名来自 28 个国家的 16-74 岁在线成年人。来源：Ipsos, “AI Monitor 2024”（6/24），引自 Nestor Maslej et al., 'The AI Index 2025 Annual Report,' AI Index Steering Committee, Stanford HAI (4/25)



- 1 Seem Like Change Happening Faster Than Ever?  
Yes, It Is
- 2 AI User + Usage + CapEx Growth =  
Unprecedented
- 3 AI Model Compute Costs High / Rising + Inference Costs Per Token Falling =  
Performance Converging + Developer Usage Rising
- 4 AI Usage + Cost + Loss Growth =  
Unprecedented
- 5 AI Monetization Threats =  
Rising Competition + Open-Source Momentum + China's Rise
- 6 AI & Physical World Ramps =  
Fast + Data-Driven
- 7 Global Internet User Ramps Powered by AI from Get-Go =  
Growth We Have Not Seen Likes of Before
- 8 AI & Work Evolution =  
Real + Rapid

- 1 变化似乎比以往任何时候都快？是的，的确如此
- 2 AI 用户 + 使用量 + 资本支出增长 = 前所未有
- 3 AI 模型计算成本高 / 上升 + 每次 Token 的推理成本下降 = 性能趋同 + 开发者使用量上升
- 4 AI 使用量 + 成本 + 损失增长 = 前所未有
- 5 人工智能货币化威胁 = 竞争加剧 + 开源势头 + 中国的崛起
- 6 AI 与物理世界的加速发展 = 快速 + 数据驱动
- 7 Global Internet User Ramps Powered by AI from Get-Go = 增长前所未见
- 8 AI 与工作变革 = 真实的 + 快速的

For the most part, we have focused on AI momentum and monetization of desktop / mobile software... AI momentum and monetization in our physical world is, in some respects, even more head-turning. We are entering an era where intelligence is not just embedded in digital applications, but also in vehicles, machines, and defense systems. Beyond the rise of digital agents, the world is increasingly experiencing the rise of physical agents.

Self-driving fleets like Waymo’s and Tesla’s Full Self-Driving (FSD) beta are no longer science projects confined to test tracks – they’re revenue-generating deployments, logging millions of driverless miles with increasingly autonomous software loops. The stack beneath them is getting smarter, and the data is more vast and richer. Applied Intuition, for example, is building simulation platforms and software-defined vehicle systems that abstract autonomy away from hardware – so manufacturers can ship intelligence as easily as parts. Per Uber CEO Dara Khosrowshahi, *Fast forward 15, 20 years, I think that the autonomous driver is going to be a better driver than the human driver. They will have trained on lifetimes of driving that no person can, they’re not going to be distracted.*

We are seeing the early architecture of AI-native infrastructure for the physical world. In defense, companies like Anduril are redefining what defense looks like – shipping autonomous drones and counter-intrusion systems with AI in every edge node, not just the command center. In agriculture, companies like Carbon Robotics are putting AI into the dirt – using computer vision to eliminate weeds without herbicides.

We believe that these are examples of a broader shift: a world where AI turns capital assets into software endpoints. Intelligence, once confined to screens and dashboards, becomes kinetic.

在很大程度上，我们一直专注于人工智能的势头以及桌面 / 移动软件的货币化 …… 在某种程度上，人工智能在我们物理世界中的势头和货币化甚至更加引人注目。我们正在进入一个智能不仅嵌入在数字应用程序中，而且还嵌入在车辆、机器和防御系统中的时代。除了数字代理的兴起之外，世界也日益体验到物理代理的兴起。

像 Waymo 和 Tesla 的 Full Self-driving 车队 -Driving (FSD) beta 不再是局限于测试轨道的科学项目 —— 它们是产生收入的部署，通过日益自主的软件循环记录数百万英里的无人驾驶里程。它们下面的堆栈变得越来越智能，数据也更加庞大和丰富。例如，Applied Intuition 正在构建模拟平台和软件定义的车辆系统，将自主性从硬件中抽象出来 —— 因此制造商可以像运输零件一样轻松地运输智能。根据 Uber 首席执行官 Dara Khosrowshahi 的说法，快进 15、20 年，我认为自动驾驶员将比人类驾驶员更好。他们将接受人一生都无法企及的驾驶训练，他们不会分心。

我们正在看到物理世界中 AI 原生基础设施的早期架构。在国防领域，像 Anduril 这样的公司正在重新定义国防的形态 – 在每个边缘节点（而不仅仅是指挥中心）中，都通过 AI 来运输自主无人机和反入侵系统。在农业领域，像 Carbon Robotics 这样的公司正在将 AI 应用于土壤 – 使用计算机视觉来消除杂草，而无需使用除草剂。

我们认为这些都是更广泛转变的例子：一个 AI 将资本资产转化为软件端点的世界。曾经局限于屏幕和仪表板的智能，正在变得具有动能。

# Physical World AI – Vertically-Integrated Electric Vehicles (Tesla) = ~100x Increase in Fully Self-Driven Miles Over Thirty-Three Months

# 物理世界 AI – 垂直整合电动汽车 (Tesla) =~100 在三十三个月内完全自动驾驶里程增加 x 倍

## Tesla Vertically-Integrated Electric Vehicles

## Tesla 垂直整合电动汽车



*For full self-driving, we've released version 12, which is a complete architectural rewrite compared to prior versions. This is end-to-end artificial intelligence...*

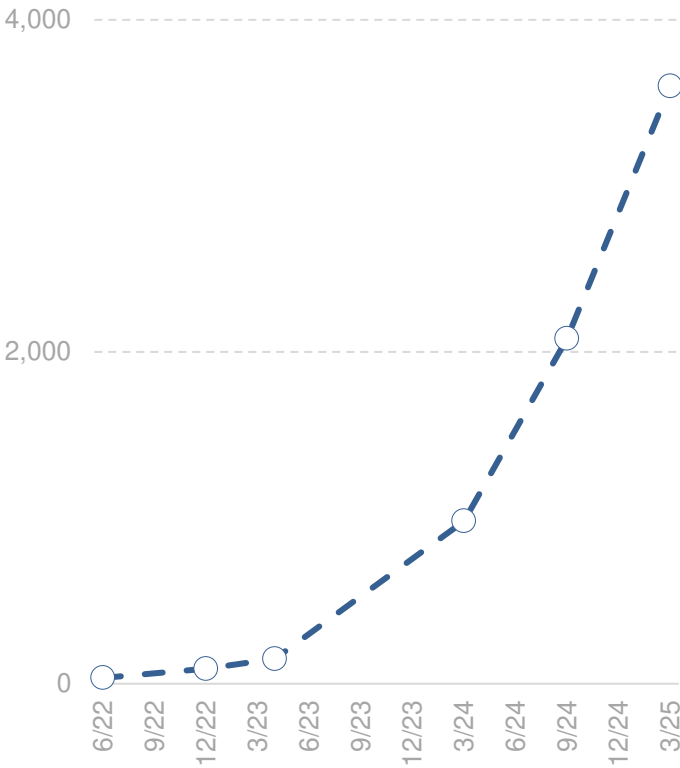
*...And it really is...quite a profound difference...*

*...So, this is the first time AI is being used, not just for object perception, but for path planning and vehicle controls. We replaced 330,000 lines of C++ code with neural nets. It's really quite remarkable. So, as a side note, I think Tesla is probably the most probably the most efficient company in the world for AI inference. Out of necessity.*

**- Tesla CEO Elon Musk, 1/24**

Tesla Cumulative Fully Self-Driven Miles (MM) – 6/22-3/25, per Tesla

Tesla Cumulative Full Self-Driving Miles Driven, MM



对于完全自动驾驶，我们发布了版本 12，与之前的版本相比，这是一个完整的架构重写。这是端到端人工智能 .....

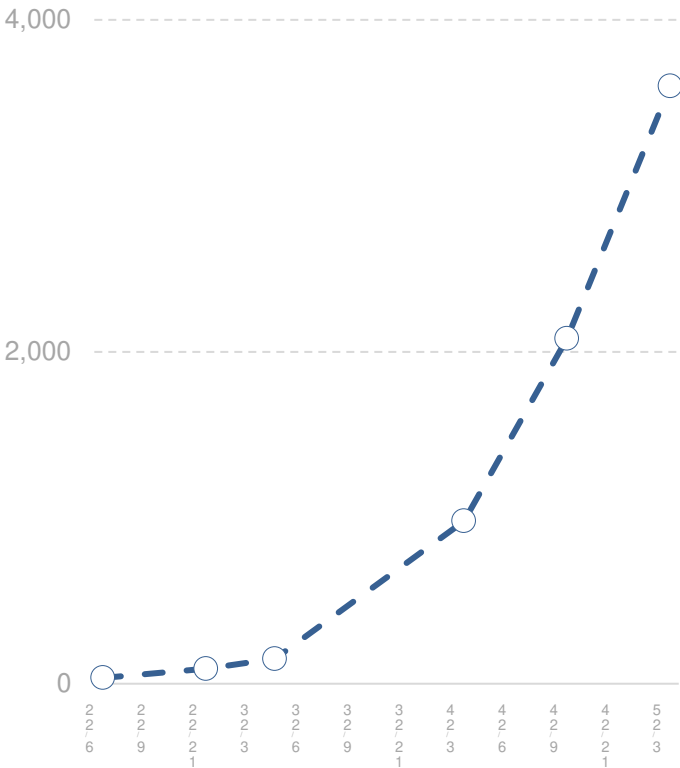
..... 而且这确实 ..... 是一个非常深刻的差异 .....

... 所以，这是人工智能首次被使用，不仅用于物体感知，还用于路径规划和车辆控制。我们用神经网络替换了 330,000 行 C++ 代码。这真的很了不起。所以，作为旁注，我认为特斯拉可能是世界上人工智能推理效率最高的公司。迫于无奈。

**- 特斯拉 CEO 埃隆·马斯克, 1/24**

特斯拉累计全自动驾驶里程 (MM) –6/22-3/25，根据特斯拉

特斯拉累计全自动驾驶里程 (MM)



Source: Tesla Disclosures & Q1:25 Investor Deck

来源: Tesla Disclosures & Q1:25 Investor Deck

# Physical World AI – Fully-Autonomous Vehicles (Waymo) = 0% to 27% Share of San Francisco Rideshares Over Twenty Months, per YipitData

# 物理世界 AI – 全自动驾驶汽车 (Waymo) =根据 YipitData, 在 20 个月内占据旧金山网约车市场份额的 0% 至 27%

## Waymo Fully-Autonomous Vehicles



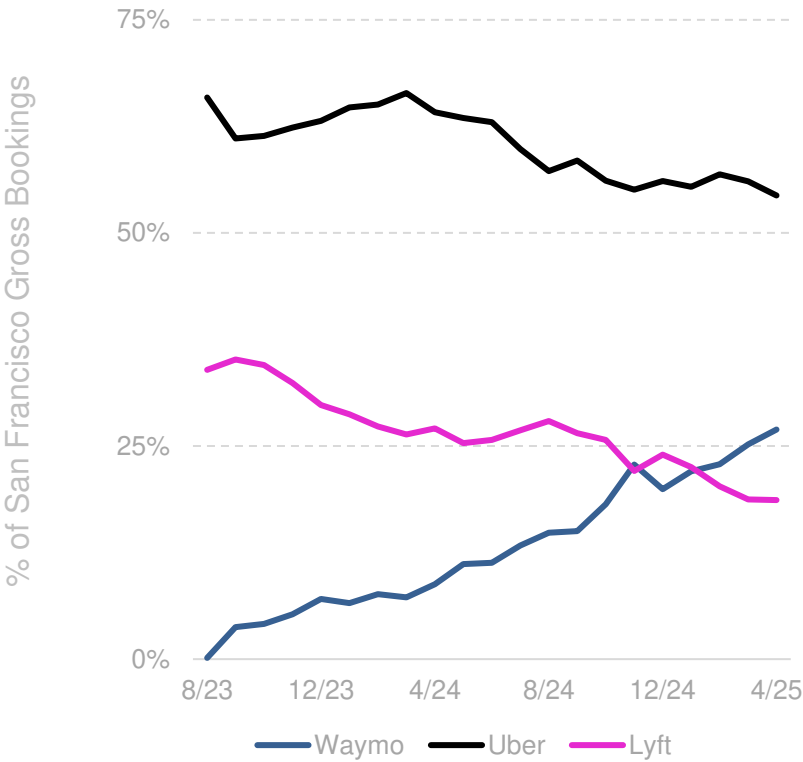
*[We are creating] an end-to-end, very, very robust, and large end-to-end system that's multi-modal in its foundation so that perception planning and prediction... can become even more robust than it is today.*

**- Waymo Co-CEO Tekedra Mawakana, 1/25**

*What we've done in San Francisco is prove to ourselves – and to the world – that not only does autonomy work, but it works at scale in a market and can be a viable commercial product.*

**- Waymo Co-CEO Dmitri Dolgov, 3/25**

Estimated Market Share (Gross Bookings) – 8/23-4/25, San Francisco Operating Zone, per YipitData



Note: Data derived from USA-user email receipt panel composed of >1mm monthly transacting USA email accounts from all available domains. Paid rides only. Numbers are estimates due to sample size. Source: Waymo, Tech Brew (1/25), Fast Company (3/25), YipitData (4/4/25)

## Waymo 全自动驾驶汽车



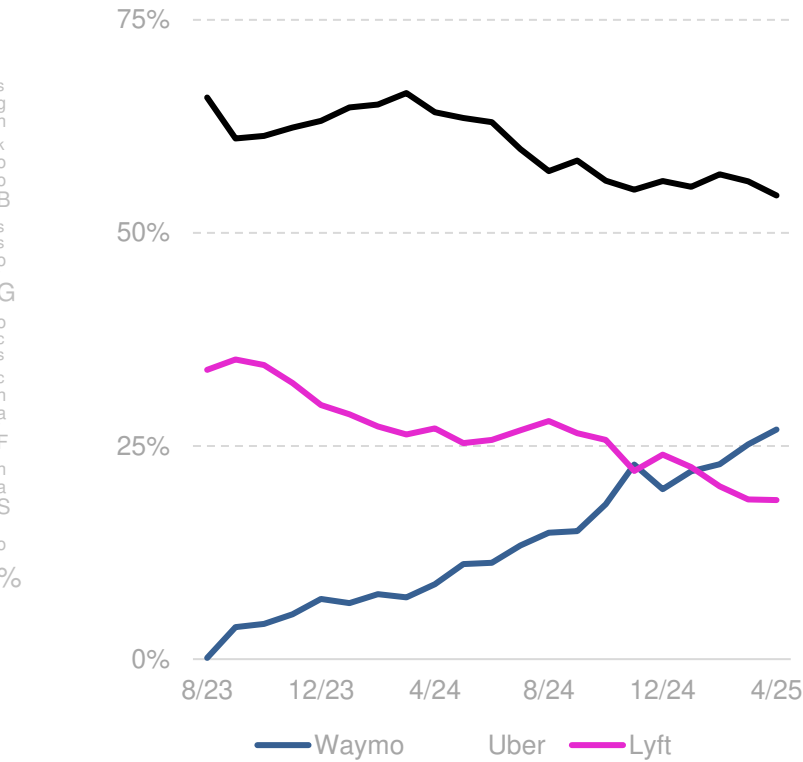
*[我们正在创建一个] 端到端、非常非常稳健且大型的端到端系统，该系统在其基础上是多模式的，以便感知规划和预测 …… 能够变得比今天更加稳健。*

**- Waymo 联合首席执行官 Tekedra Mawakana, 1/25**

*我们在旧金山所做的事情向我们自己证明了 —— 也向世界证明了 —— 自主不仅有效，而且可以在市场上大规模运行，并且可以成为可行的商业产品。*

**- Waymo Co-CEO Dmitri Dolgov, 3/25**

估计市场份额（总预订量）– 8/23-4/25，旧金山运营区，根据 YipitData

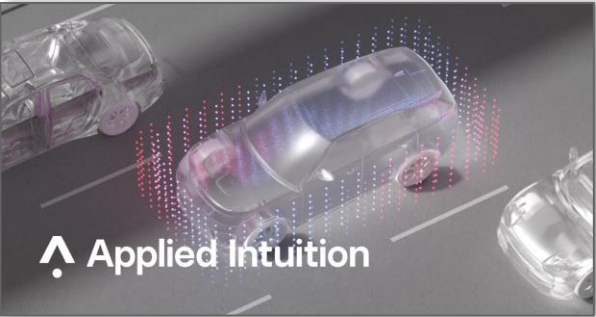


注意：数据来源于美国用户电子邮件收据面板，该面板由来自所有可用域的 >1 百万月度交易美国电子邮件帐户组成。仅限付费乘车。由于样本量，数字为估计值。来源：Waymo, Tech Brew (1/25), Fast Company (3/25), YipitData (4/4/25)

# Physical World AI –Vehicle Intelligence (Applied Intuition) = Serving Automotive, Trucking, Construction & Defense

# 物理世界 AI – 车辆智能 (Applied Intuition) = 服务于汽车、卡车运输、建筑和国防

## Applied Intuition Vehicle Intelligence



*Within the last few years, we've seen massive advances in artificial intelligence that will have groundbreaking impacts on the industries that Applied Intuition serves. Our role as a leader in the ecosystem is to bring the best of what Silicon Valley has to offer to our global customer base.*

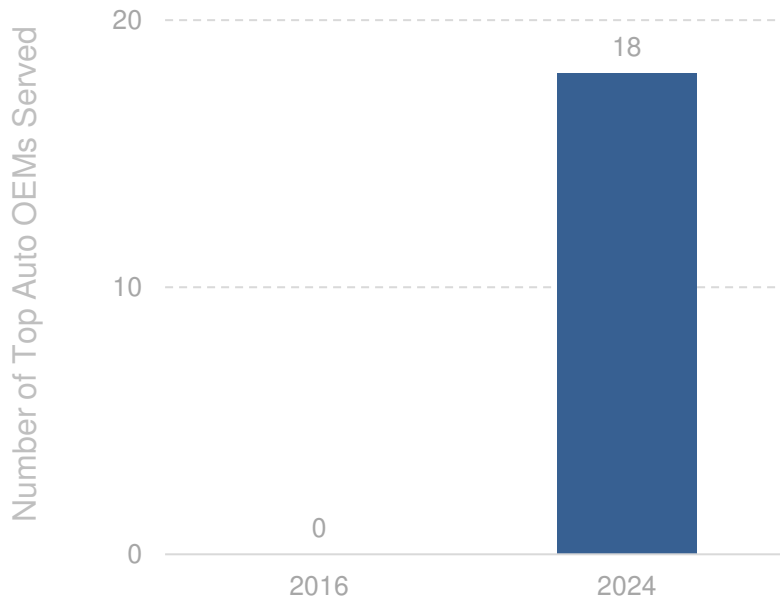
**- Applied Intuition Co-Founder & CEO Qasar Younis, 3/24**

*We've seen accelerating adoption of our AI-powered tools, autonomy software, and vehicle operating system as traditional OEMs are seeing strong ROI. The Defense sector is also looking for vehicle intelligence solutions. We've provided our off-road autonomy stack for defense for several years, and have expanded our defense tech product portfolio significantly over the past year.*

**- Applied Intuition Co-Founder & CTO Peter Ludwig, 5/25**

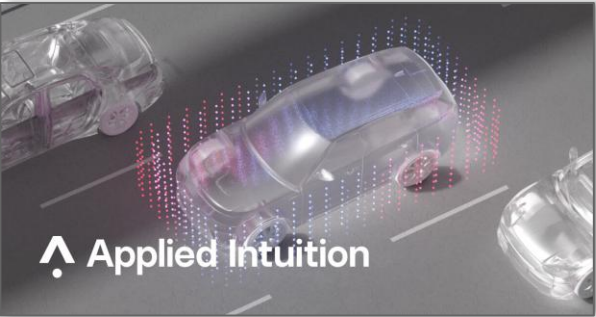
Note: OEM = Original Equipment Manufacturer.  
Source: Applied Intuition

Applied Intuition Top Global Auto OEMs Served – 2016-2024, per Applied Intuition\*



*\* Applied Intuition serves a broad base of customers in different verticals, such as Porsche / Toyota (auto), Traton / Isuzu (trucking), Caterpillar (construction) and several US military branches (defense).*

## Applied Intuition 车辆智能



在过去几年中，我们看到了人工智能的巨大进步，这将对 Applied Intuition 服务的行业产生突破性的影响。我们作为生态系统领导者的角色是将硅谷最好的东西带给我们的全球客户群。

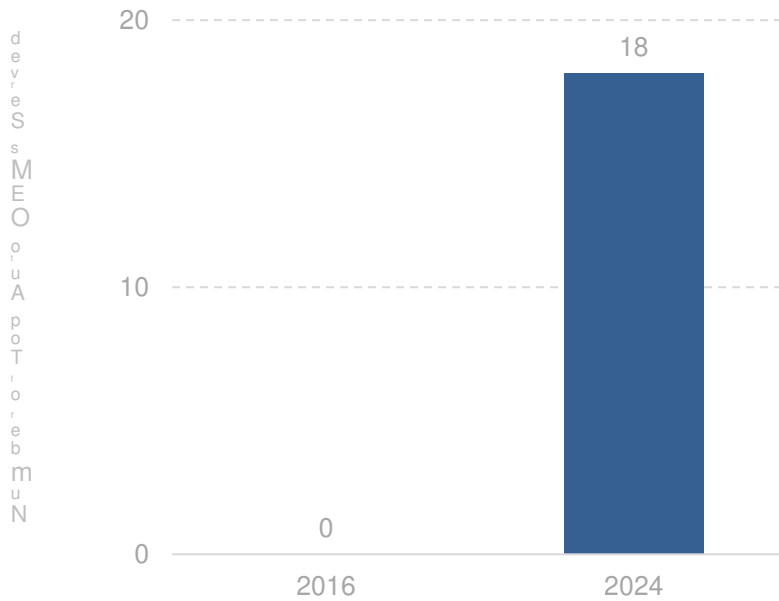
- Applied Intuition 联合创始人兼首席执行官 Qasar Younis, 3/24

我们已经看到，随着传统原始设备制造商看到强劲的投资回报率，我们的人工智能驱动工具、自主软件和车辆操作系统正在加速应用。国防部门也在寻找车辆智能解决方案。多年来，我们一直为国防部门提供越野自主堆栈，并在过去一年中大幅扩展了我们的国防技术产品组合。

- Applied Intuition 联合创始人兼首席技术官 Peter Ludwig, 5/25

注意：OEM = 原始设备制造商。来源：Applied Intuition

Applied Intuition 服务的顶级全球汽车 OEM – 2016-2024，数据来源：Applied Intuition\*



*\* Applied Intuition 为不同垂直领域的广泛客户群提供服务，例如保时捷 / 丰田（汽车）、Traton/ 五十铃（卡车运输）、卡特彼勒（建筑）和多个美国军事部门（国防）。*



# Physical World AI – USA Defense (Anduril) = +2x Y/Y Revenue Growth for Last Two Years

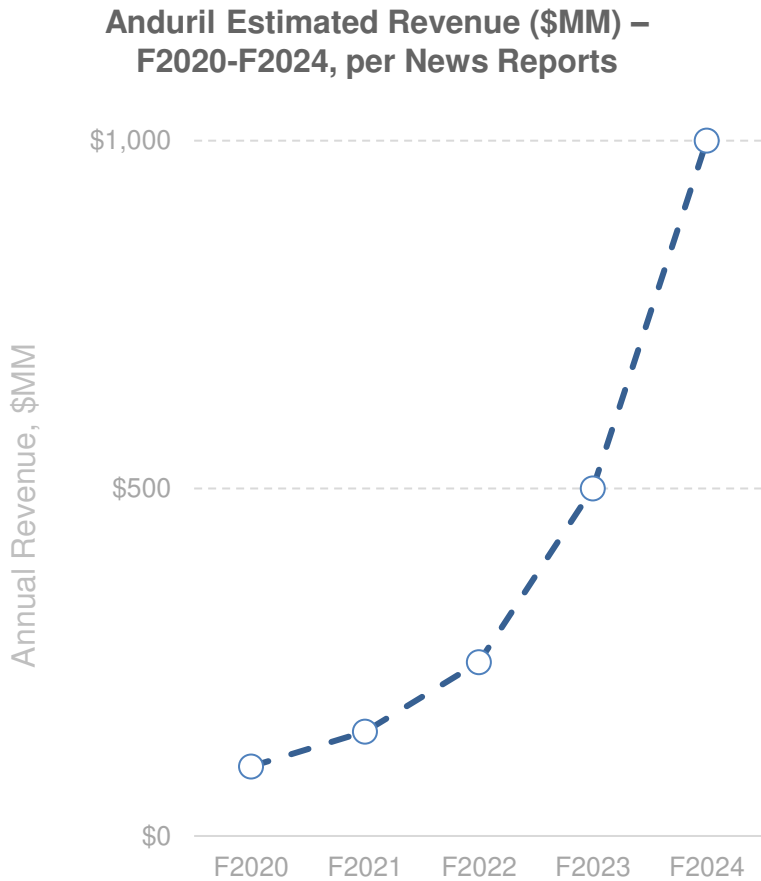
## Anduril AI-Enabled Autonomous USA Defense Systems



*At Anduril, we firmly believe that today’s most pressing national security challenges cannot be solved without AI-enabled systems and autonomy at scale. These systems will help to keep our service members safe and empower them to make better decisions at the speed of modern warfare...*

*...When developed and deployed properly, [AI and autonomous systems] can make warfare more proportional, more precise, and less indiscriminate than it has ever been before.*

**- Anduril Co-Founder & CEO Brian Schimpf, 12/23**



Source: Anduril, Forbes, TechCrunch, CNBC

# 物理世界 AI – USA Defense (Anduril) =+2x 过去两年的同比增长率

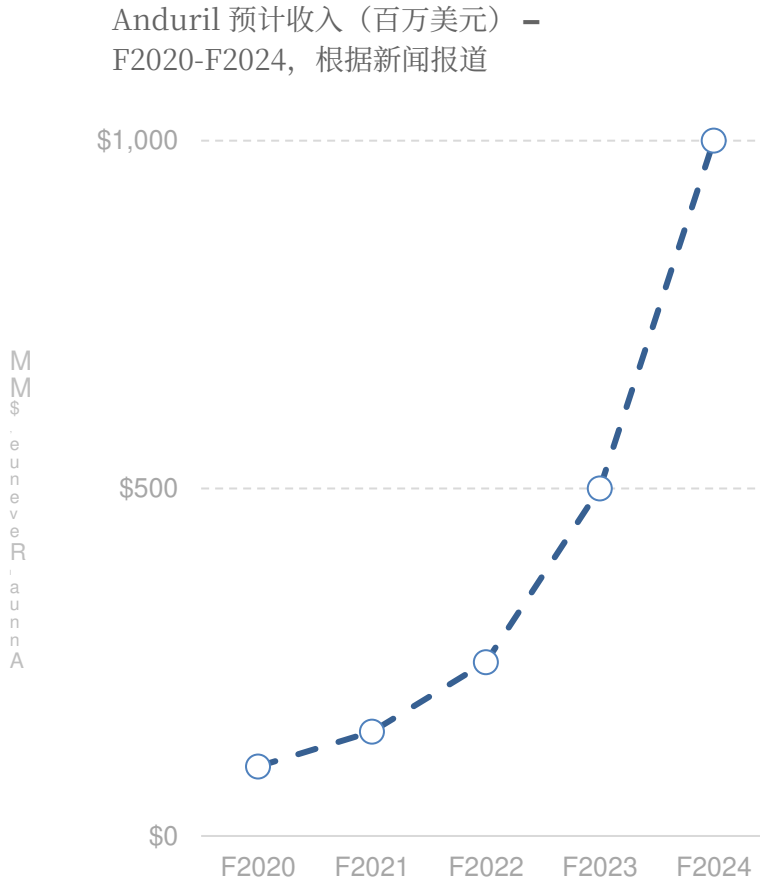
## Anduril AI 支持的自主 USA 防御系统



在 Anduril，我们坚信，如果没有 AI 支持的系统和大规模的自主性，就无法解决当今最紧迫的国家安全挑战。这些系统将有助于保护我们的军人安全，并使他们能够在现代战争的速度下做出更好的决策……

… 如果开发和部署得当，[AI 和自主系统] 可以使战争比以往任何时候都更加均衡、精确和更少滥杀无辜。

**- Anduril 联合创始人兼 CEO Brian Schimpf, 12/23**



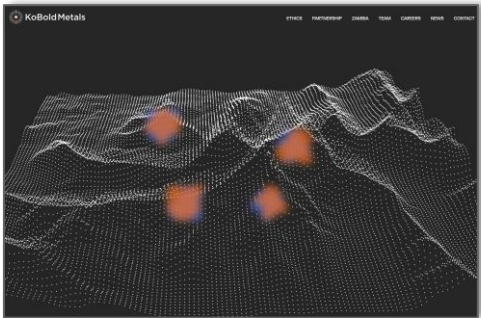
Source: Anduril, Forbes, TechCrunch, CNBC

# Physical World AI = AI-Driven Mining Exploration (KoBold Metals) = Reversing Trend in Exploration Inefficiency

# 物理世界 AI = AI 驱动的采矿勘探 (KoBold Metals) = 扭转勘探效率低下的趋势

## KoBold Metals AI-Driven Mining Exploration

## KoBold Metals AI 驱动的采矿勘探



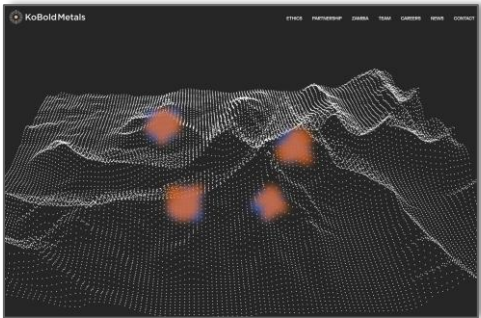
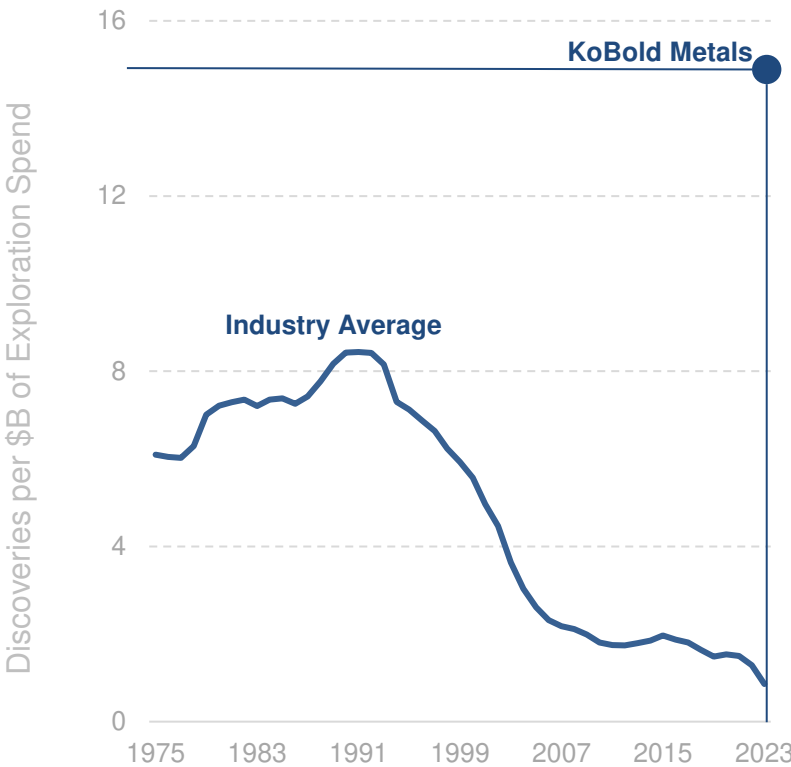
*We're looking to expand and diversify the supply of these metals all over the world, but we're taking a totally different approach [from conventional mining companies]. Two-thirds of our team are software engineers or data scientists.*

**- KoBold Metals Co-Founder & CEO Kurt House, 12/22**

*KoBold's Machine Prospector technology combines never before used datasets with conventional geochemical, geophysical, & geological data in statistical association models to identify prospects. KoBold's technology accelerates exploration by efficiently screening large regions & makes our search more effective by identifying the most promising locations.*

**- KoBold Metals Website**

Mineral Deposit Discoveries per \$B of Exploration Spend – 1975-2023, per KoBold Metals



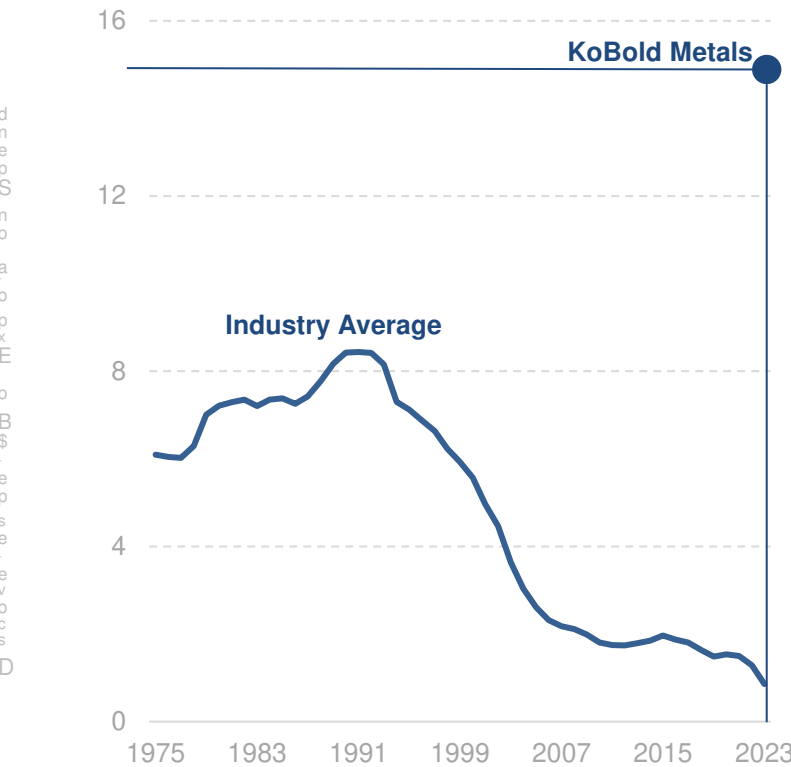
我们希望扩大并多样化全球这些金属的供应，但我们采取了一种与传统矿业公司 [完全不同的方法]。我们团队的三分之二是软件工程师或数据科学家。

- KoBold Metals 联合创始人兼首席执行官 Kurt House, 12/22

KoBold 的 Machine Prospector 技术将前所未有的数据集与传统的地球化学、地球物理和地质数据结合在统计关联模型中，以识别前景。KoBold 的技术通过有效筛选大区域来加速勘探，并通过识别最有希望的地点使我们的搜索更有效。

**- KoBold Metals Website**

Mineral Deposit Discoveries per \$B of Exploration Spend – 1975-2023, per KoBold Metals





Carbon Robotics AI-Driven Agricultural Modernization

Carbon Robotics AI 驱动 的农业现代化



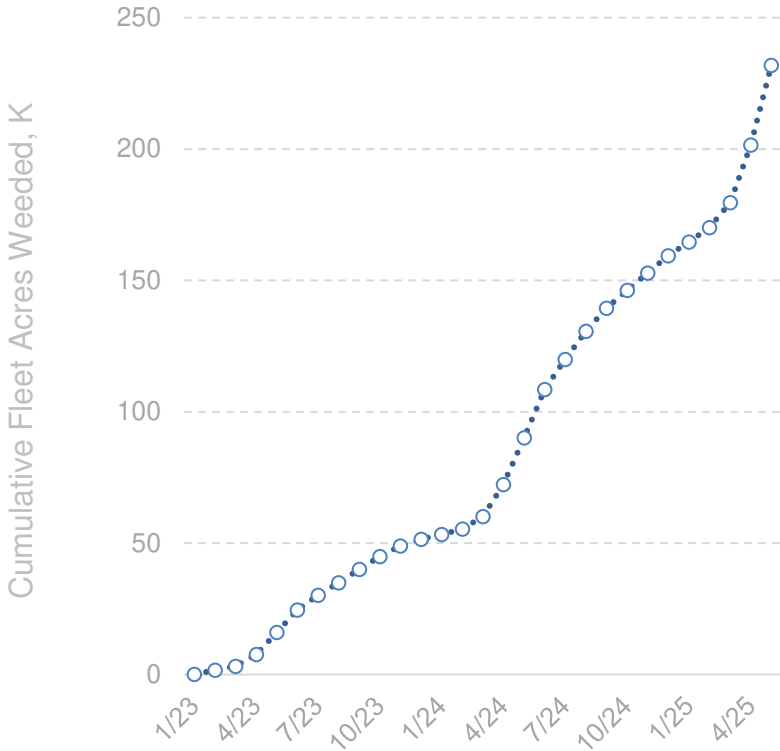
*The LaserWeeder leverages our sophisticated laserweeding technology, driven by AI deep learning models and computer vision software, to efficiently identify, target, and eliminate weeds by zapping them at the meristem. The implement can cover up to 2 acres per hour and shoot up to 200,000 weeds.*

**- Carbon Robotics Founder & CEO Paul Mikesell, 12/22**

*We learned from farmers that their biggest challenges continue to be around labor and labor availability. If they could, they would run everything 24/7. They would run everything every minute of farming season to get as much done as possible.*

**- Carbon Robotics Founder & CEO Paul Mikesell, 3/25**

Carbon Robotics Cumulative Fleet Acres Weeded (K) – 1/23-5/25, per Carbon Robotics



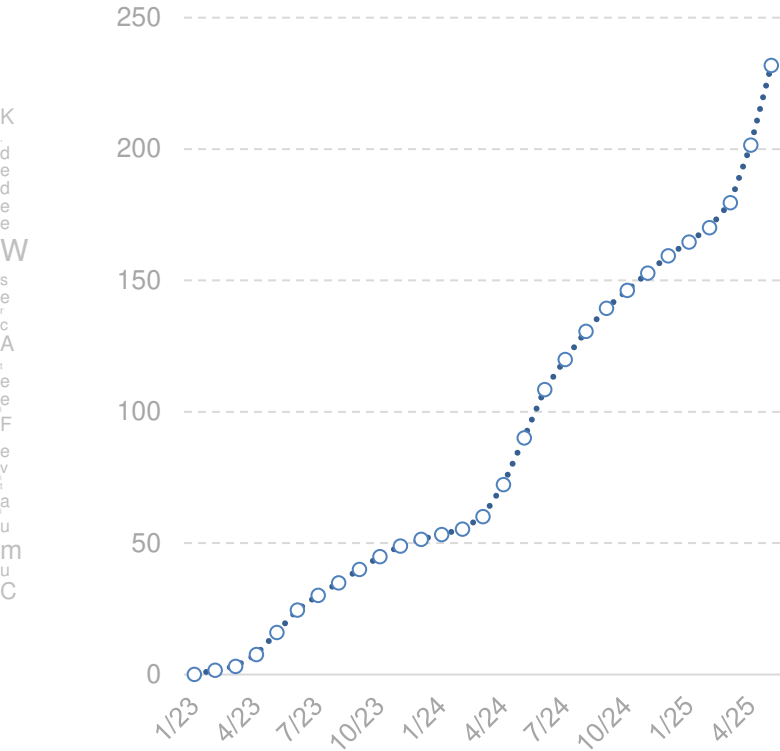
LaserWeeder 利用我们先进的激光除草技术，由 AI 深度学习模型和计算机视觉软件驱动，通过在分生组织处对其进行照射，从而有效地识别、瞄准和消除杂草。该设备每小时可覆盖多达 2 英亩的土地，并可清除多达 200,000 株杂草。

- Carbon Robotics 创始人兼 CEO Paul Mikesell, 12/22

我们从农民那里了解到，他们面临的 最大挑战仍然是劳动力和劳动力可用性问题。如果可以的话，他们会全天候 24/7 运行所有设备。他们会在耕作季节的每一分钟都运行所有设备，以尽可能多地完成工作。

- Carbon Robotics 创始人兼首席执行官 Paul Mikesell, 3/25

Carbon Robotics 累计车队除草面积（千英亩） – 1/23-5/25，数据来源：Carbon Robotics



Halter AI-Driven Intelligent Grazing

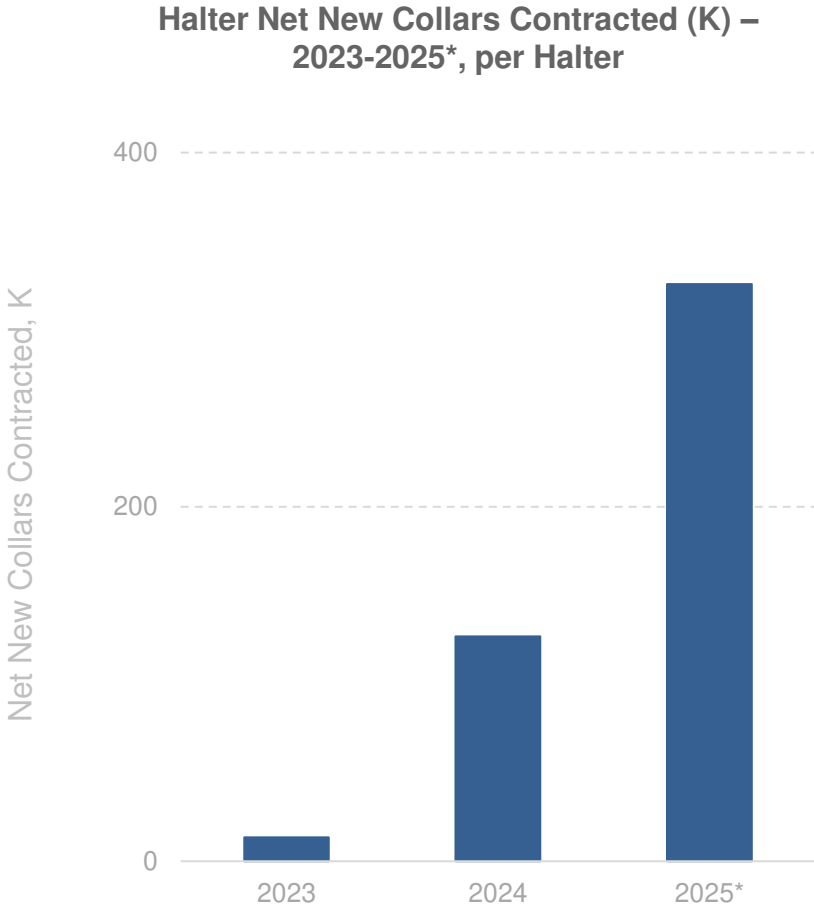
Halter AI 驱动的智能放牧



*We’ve seen firsthand the care and dedication ranchers have for their land and animals. We’ve also seen how agriculture, one of the oldest and most vital industries, has yet to receive the full benefits of modern technology. This leaves enormous opportunity for ranchers to unlock greater productivity and sustainability across their operations.*

*We believe grazing management holds the key. Effective rotational grazing enables more efficient use of natural resources and increased productivity, while also enhancing soil health and improving root structures to sequester more carbon. We don’t believe more productivity needs to come at the cost of sustainability. We can do good for ranchers, and the planet.*

- Halter (as of 5/25)

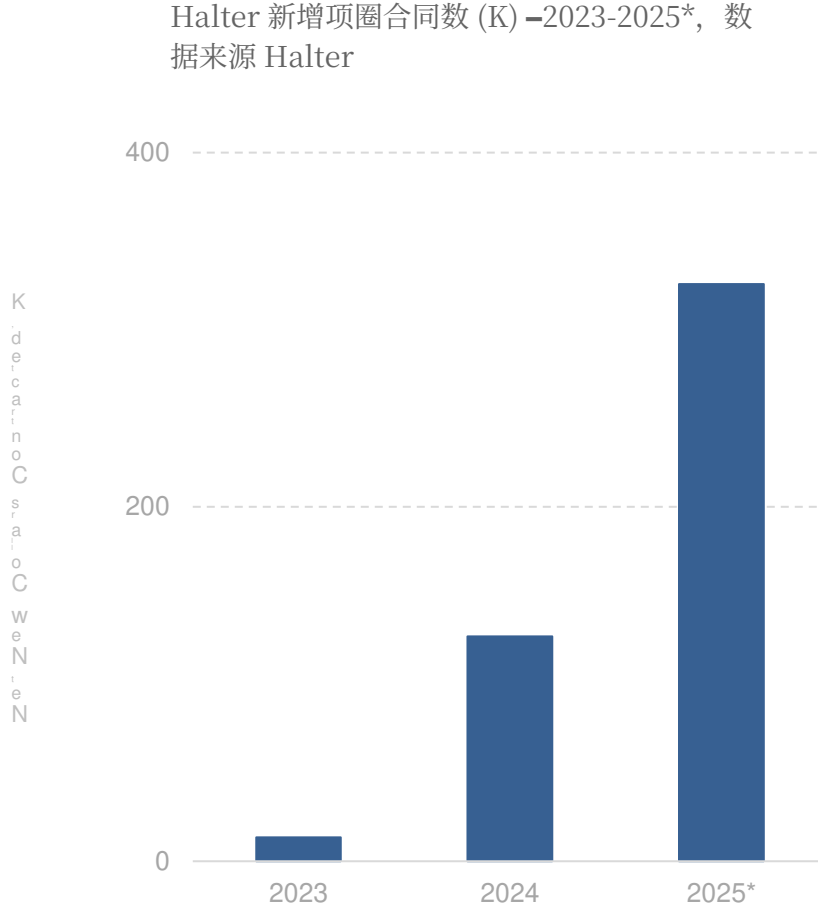


\*2025 figures annualized as of Q1:25. Source: Halter (5/25)

我们亲眼目睹了牧场主对其土地和动物的关怀和奉献。我们还看到农业，这个最古老、最重要的行业之一，尚未充分受益于现代科技。这为牧场主释放更高的生产力和整个运营的可持续性留下了巨大的机会。

我们认为放牧管理是关键。有效的轮牧能够更有效地利用自然资源并提高生产力，同时还能增强土壤健康并改善根系结构，从而封存更多的碳。我们不认为提高生产力需要以牺牲可持续性为代价。我们可以为牧场主和地球做好事。

- Halter (as of 5/25)



\*2025 年数据已按截至 25 年第一季度进行年度化。来源：Halter (5/25)

- 1 Seem Like Change Happening Faster Than Ever?  
Yes, It Is
- 2 AI User + Usage + CapEx Growth =  
Unprecedented
- 3 AI Model Compute Costs High / Rising + Inference Costs Per Token Falling =  
Performance Converging + Developer Usage Rising
- 4 AI Usage + Cost + Loss Growth =  
Unprecedented
- 5 AI Monetization Threats =  
Rising Competition + Open-Source Momentum + China's Rise
- 6 AI & Physical World Ramps =  
Fast + Data-Driven
- 7 Global Internet User Ramps Powered by AI from Get-Go =  
Growth We Have Not Seen Likes of Before
- 8 AI & Work Evolution =  
Real + Rapid

- 1 感觉变化比以往任何时候都快？是的，的确如此
- 2 AI 用户 + 使用情况 + 资本支出增长 = 前所未有
- 3 AI 模型计算成本高 / 上升 + 每次 Token 的推断成本下降 = 性能趋同 + 开发者使用量上升
- 4 AI 使用 + 成本 + 亏损增长 = 前所未有的
- 5 AI 货币化威胁 = 日益激烈的竞争 + 开源势头 + 中国的崛起
- 6 AI 与物理世界发展 = 快速 + 数据驱动
- 7 Global Internet User Ramps Powered by AI from Get-Go = 我们从未见过的增长
- 8 AI 与工作变革 = 真实的 + 迅速的

Thanks to the rise in low-cost satellite-driven Internet connectivity / access, the potential for the 2.6B (or 32% of the world’s population) that is not online to come online is increasing. These new users will start from scratch with AI functionality. Wow!

When these new users come online, they likely won’t be met by browsers and search bars. They’ll start with AI – and in their native language. Imagine a ‘first experience’ of the internet that doesn’t involve typing a query into a search engine but instead talking to a machine that talks back. Imagine skipping the traditional application layer entirely, with an agent-driven interface managing disparate tech platforms from one place while understanding users’ local language, context, and intent. An agent-first internet experience could upend existing tech hierarchies, disintermediating dominant platforms and redistributing value.

In this model, the winners wouldn’t be those who own the app, but those who own the interface.

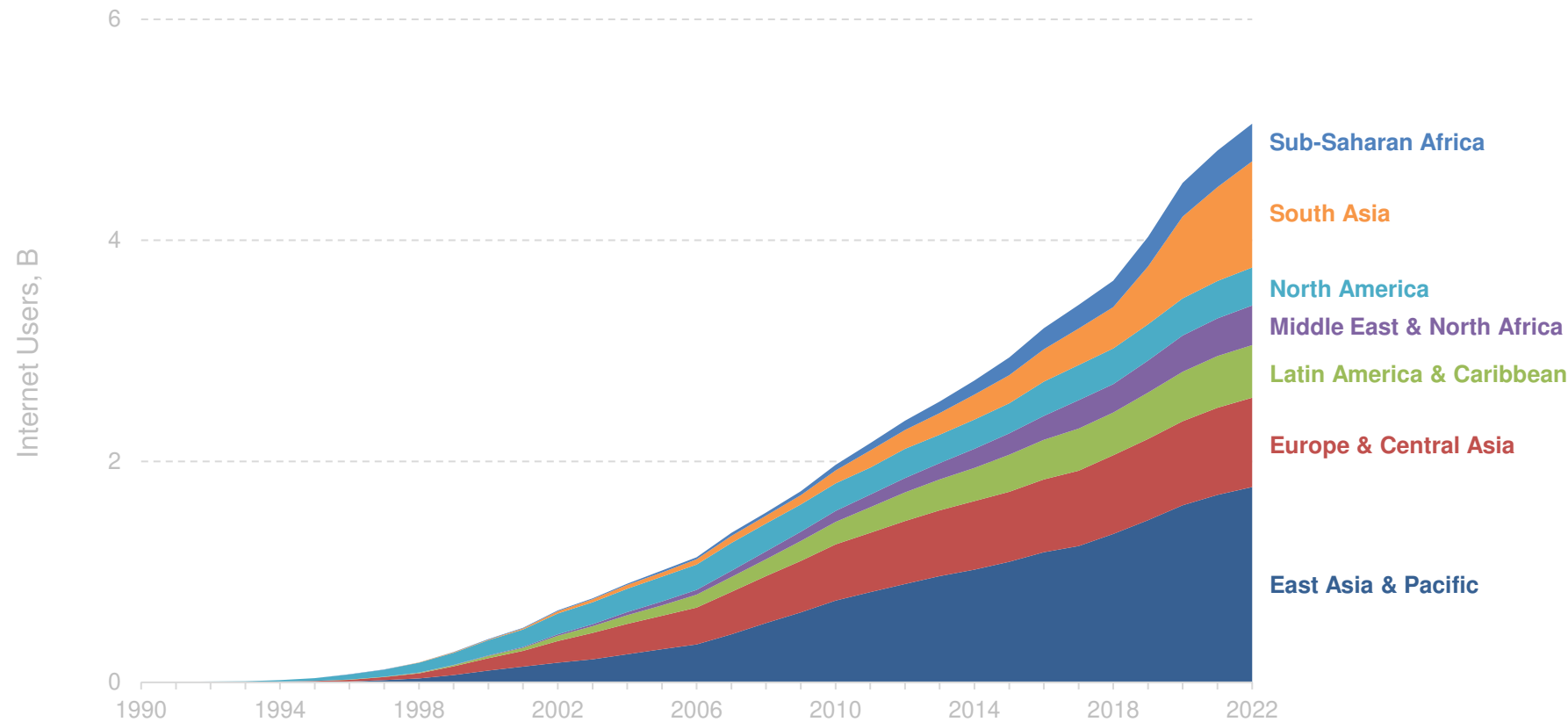
由于低成本卫星驱动的互联网连接 / 访问的兴起，未上网的 26 亿人口（占世界人口的 32%）上网的可能性越来越大。这些新用户将从零开始使用 AI 功能。哇！

当这些新用户上线时，他们可能不会遇到浏览器和搜索栏。他们将从人工智能开始 —— 并且使用他们的母语。想象一下，互联网的 “初体验” 不需要在搜索引擎中输入查询，而是与一台可以对话的机器交谈。想象一下完全跳过传统的应用层，通过一个代理驱动的界面从一个地方管理不同的技术平台，同时理解用户的本地语言、上下文和意图。一个代理优先的互联网体验可能会颠覆现有的技术等级，打破主导平台的中间地位，并重新分配价值。

在这种模式下，赢家将不是那些拥有应用程序的人，而是那些拥有界面的人。

# Global Internet Users = Epic Growth Over Past Thirty-Three Years, per ITU

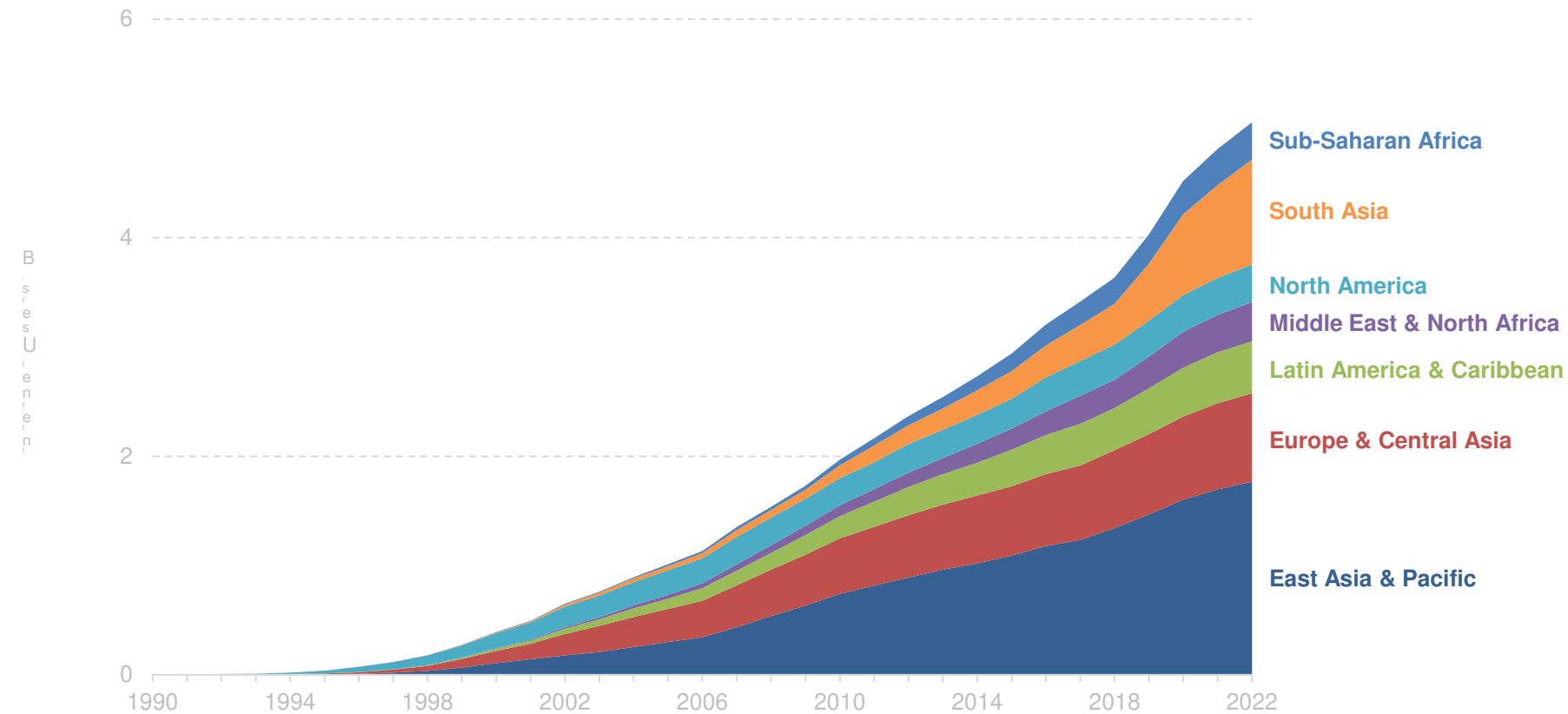
Internet Users by World Region (B) – 1990-2022, per ITU



Note: 2021 data interpolated due to data gaps for select nations. Regions are per United Nations definitions. Data is occasionally unavailable for select nations in select years, which may lead to trendline choppiness or minor discrepancies vs. global user figures. Source: United Nations / International Telecommunications Union (3/25)

# Global Internet Users = Epic Growth Over Past Thirty-Three Years, per ITU

Internet Users by World Region (B) – 1990-2022, per ITU

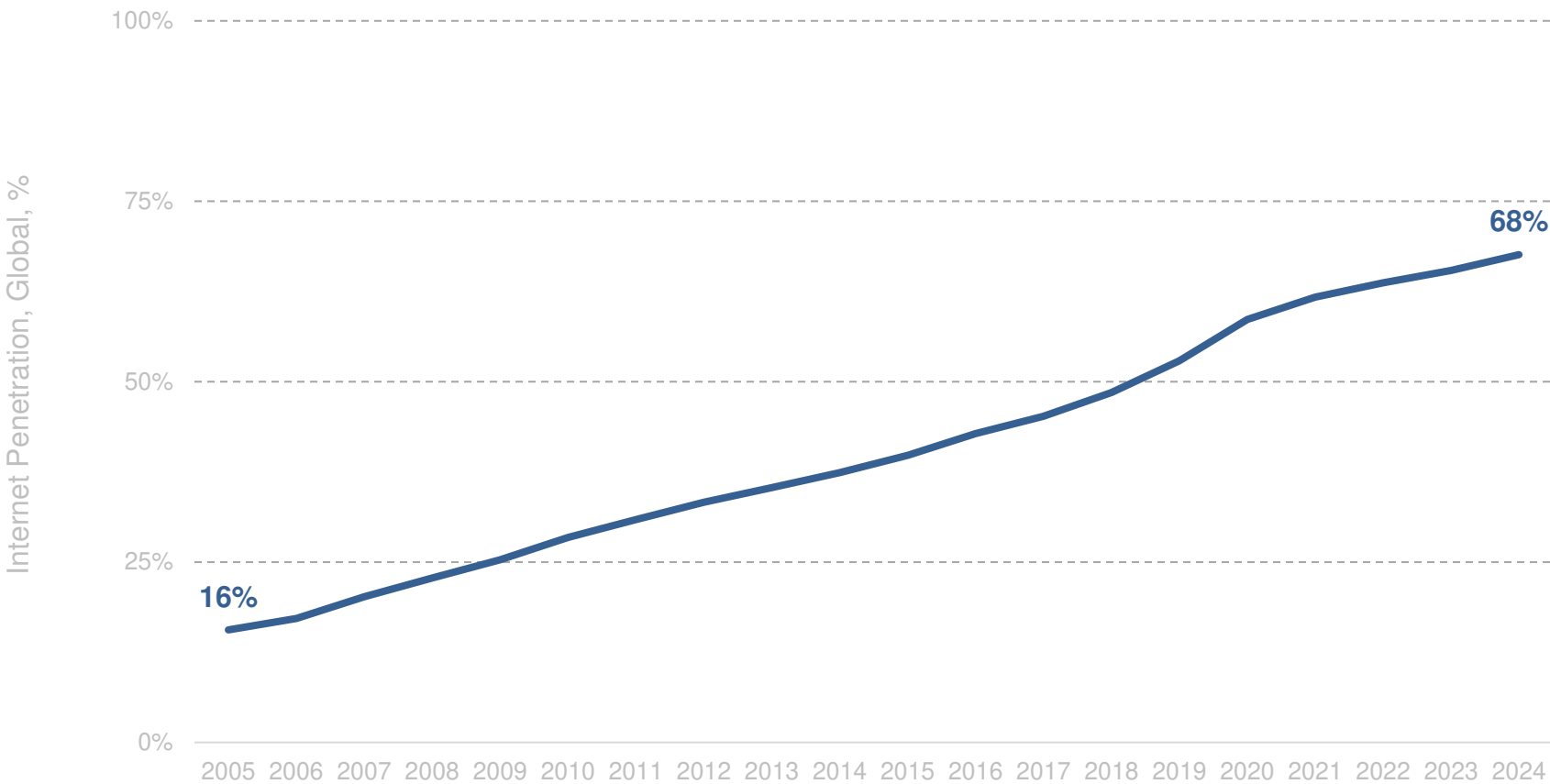


注意：由于部分国家的数据存在缺口，2021 年的数据经过插值处理。区域划分采用联合国定义。部分国家在特定年份可能缺少数据，这可能导致趋势线出现波动或与全球用户数据存在细微差异。来源：联合国 / 国际电信联盟（3/25）

Global Internet Penetration = 68% vs. 16% Nineteen Years Ago, per ITU

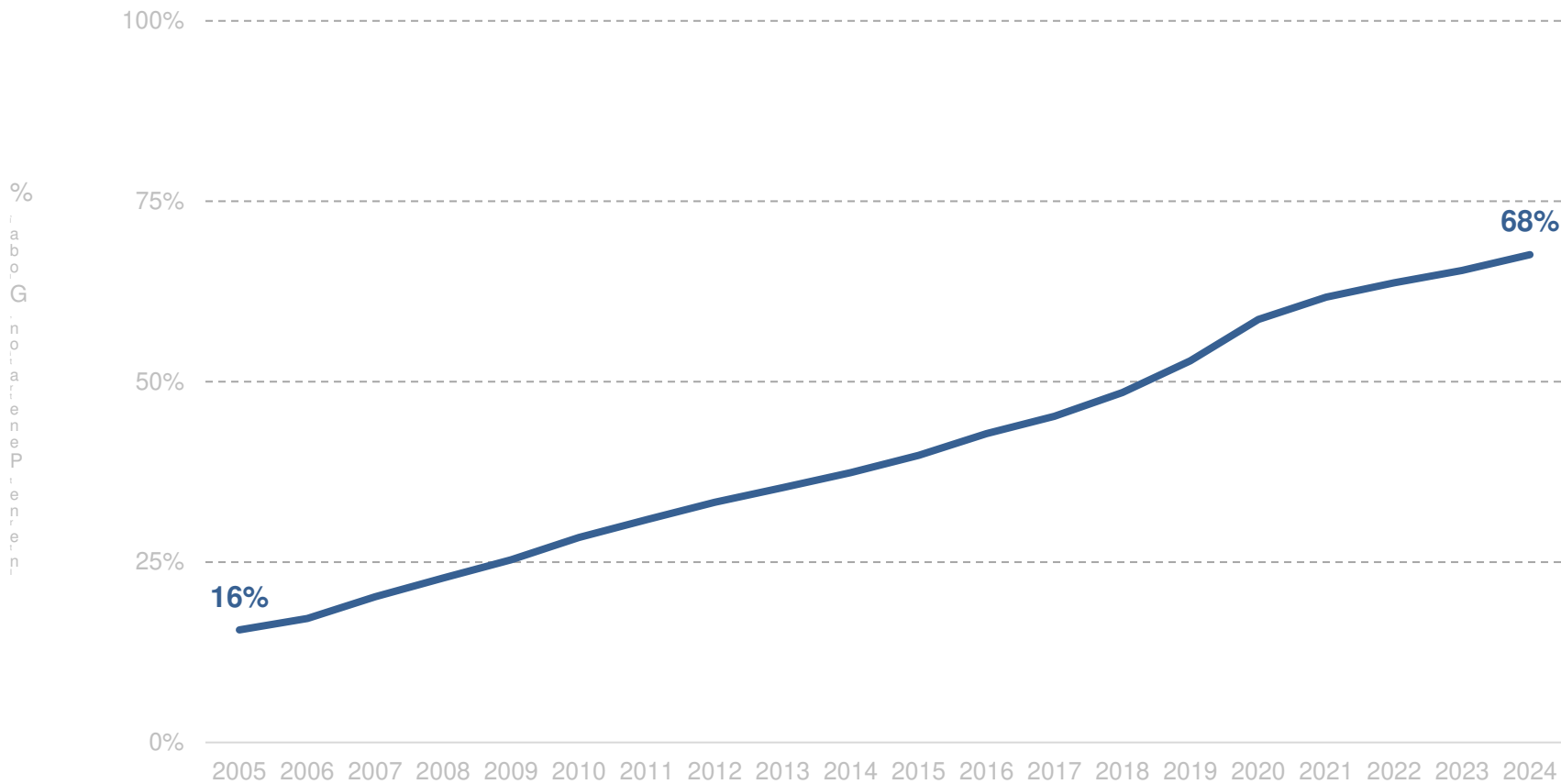
据国际电联统计，全球互联网普及率 = 为 68%，而 19 年前仅为 16%

Global Internet Penetration – 2005-2024, per ITU



Source: United Nations / International Telecommunications Union (3/25)

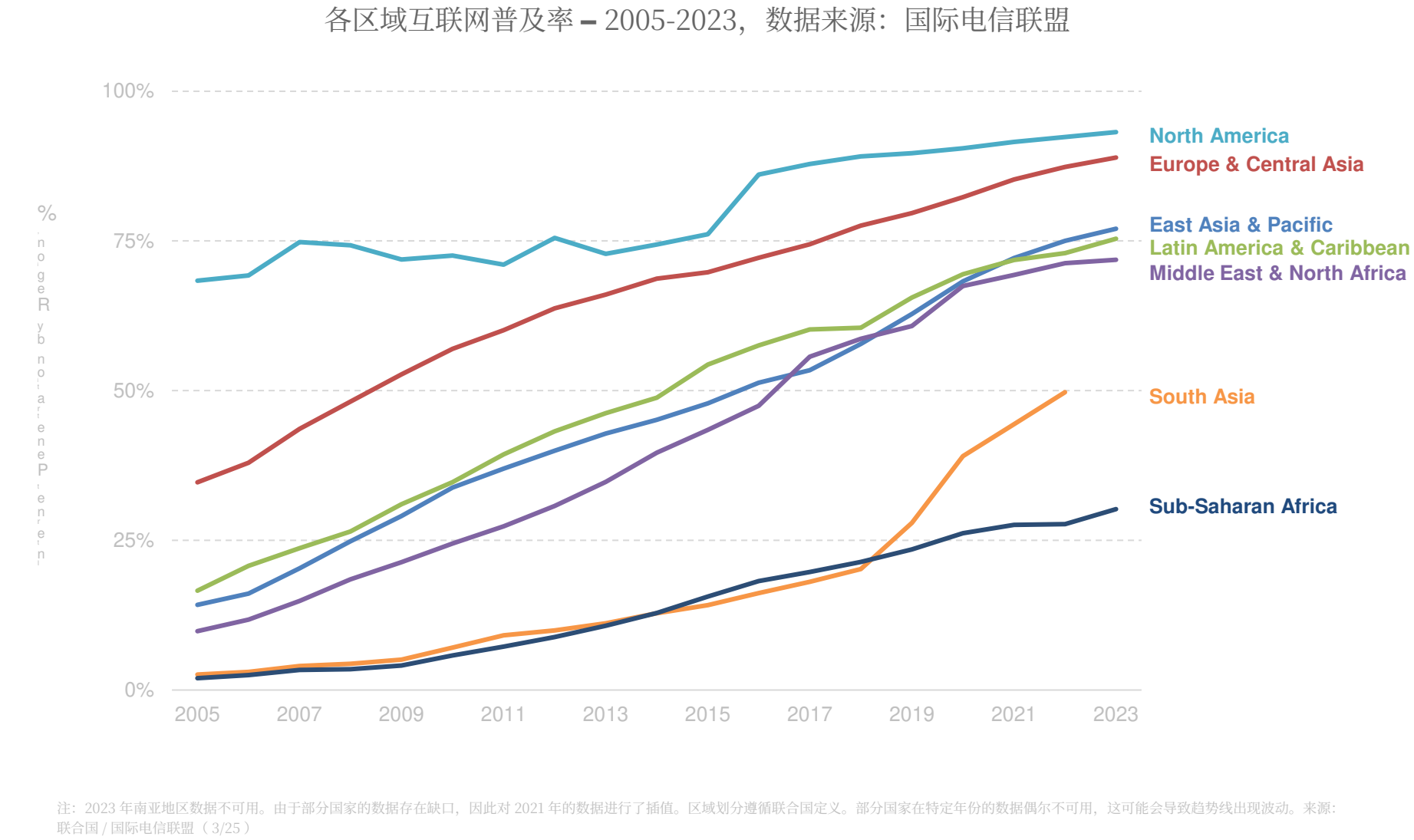
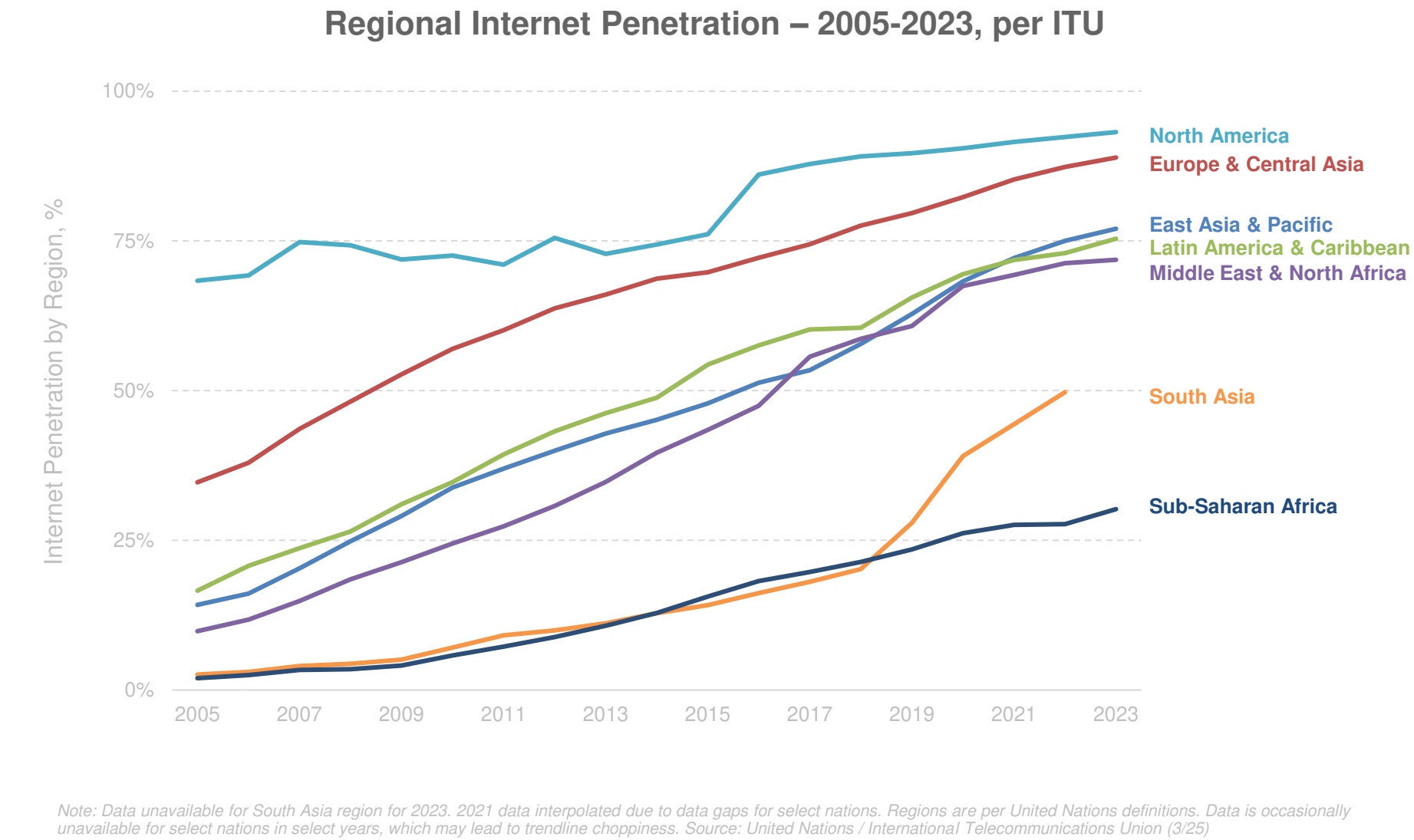
全球互联网普及率 – 2005-2024，据国际电联统计



来源：联合国 / 国际电信联盟（3/25）

Global Internet Penetration by Region @ +70% =  
All Regions Except South Asia + Sub-Saharan Africa, per ITU

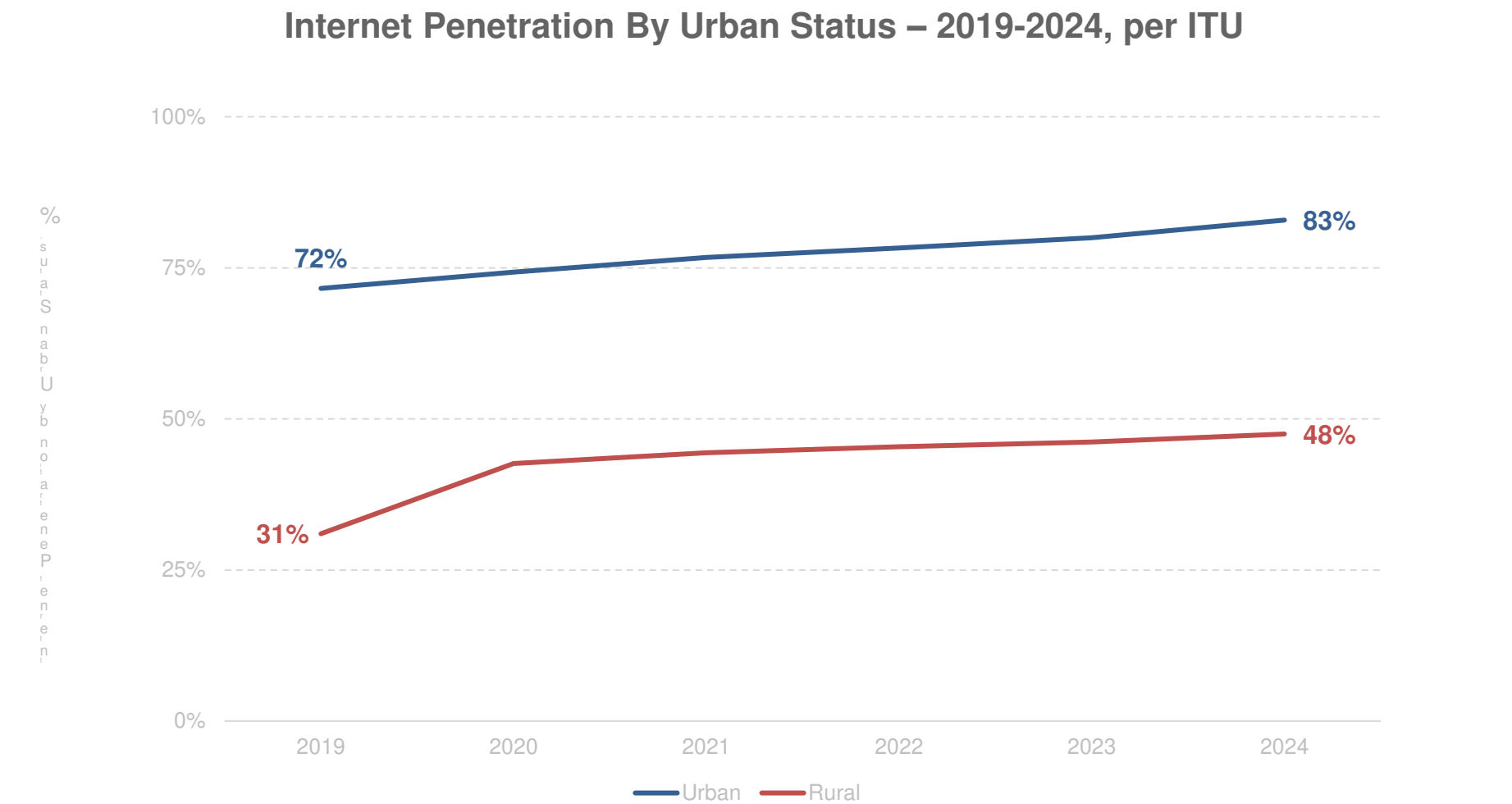
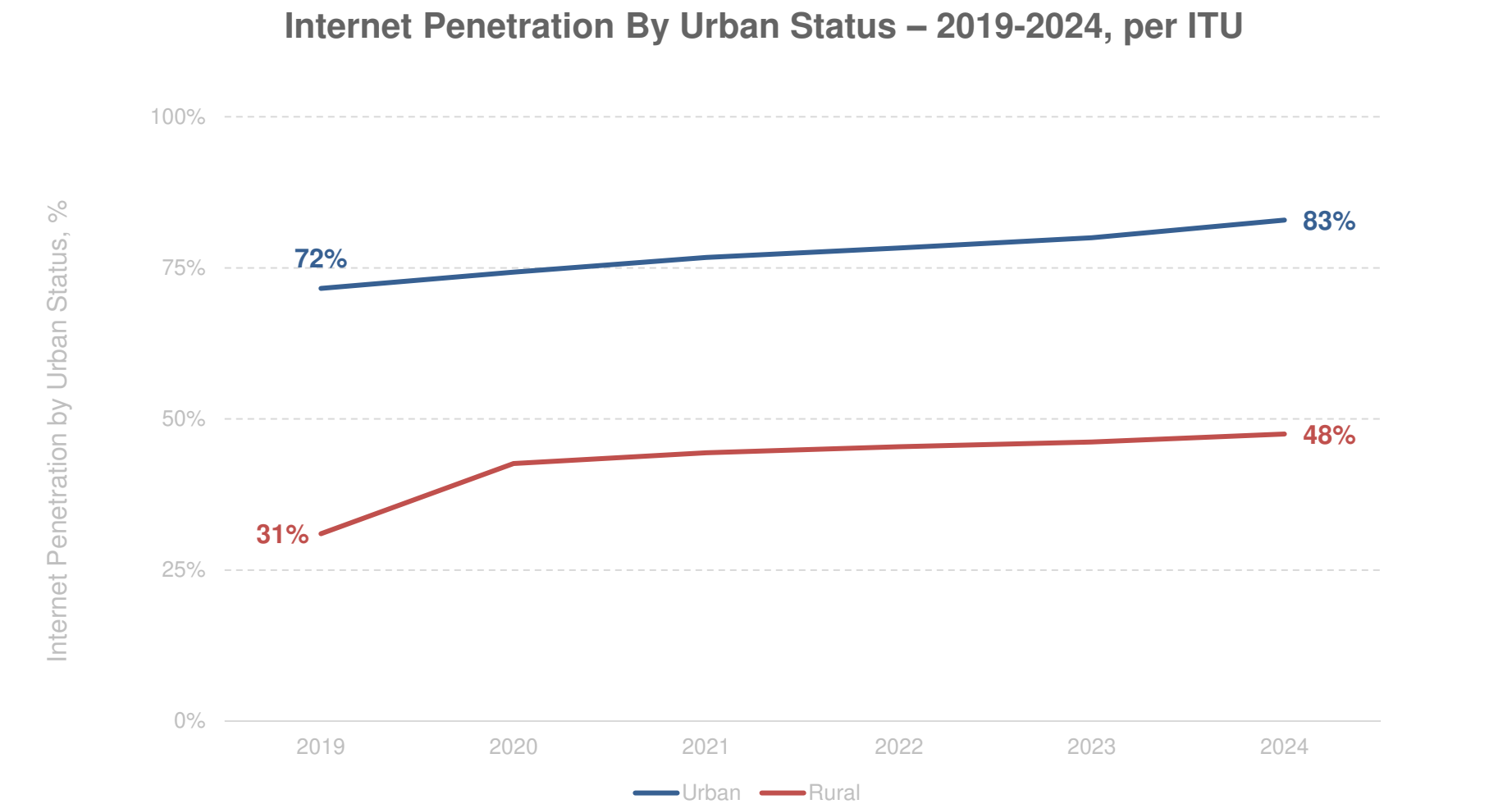
各区域互联网普及率 @ +70% = 除南亚 + 撒哈拉以南非洲之外的所有区域，数据来源：国际电信联盟





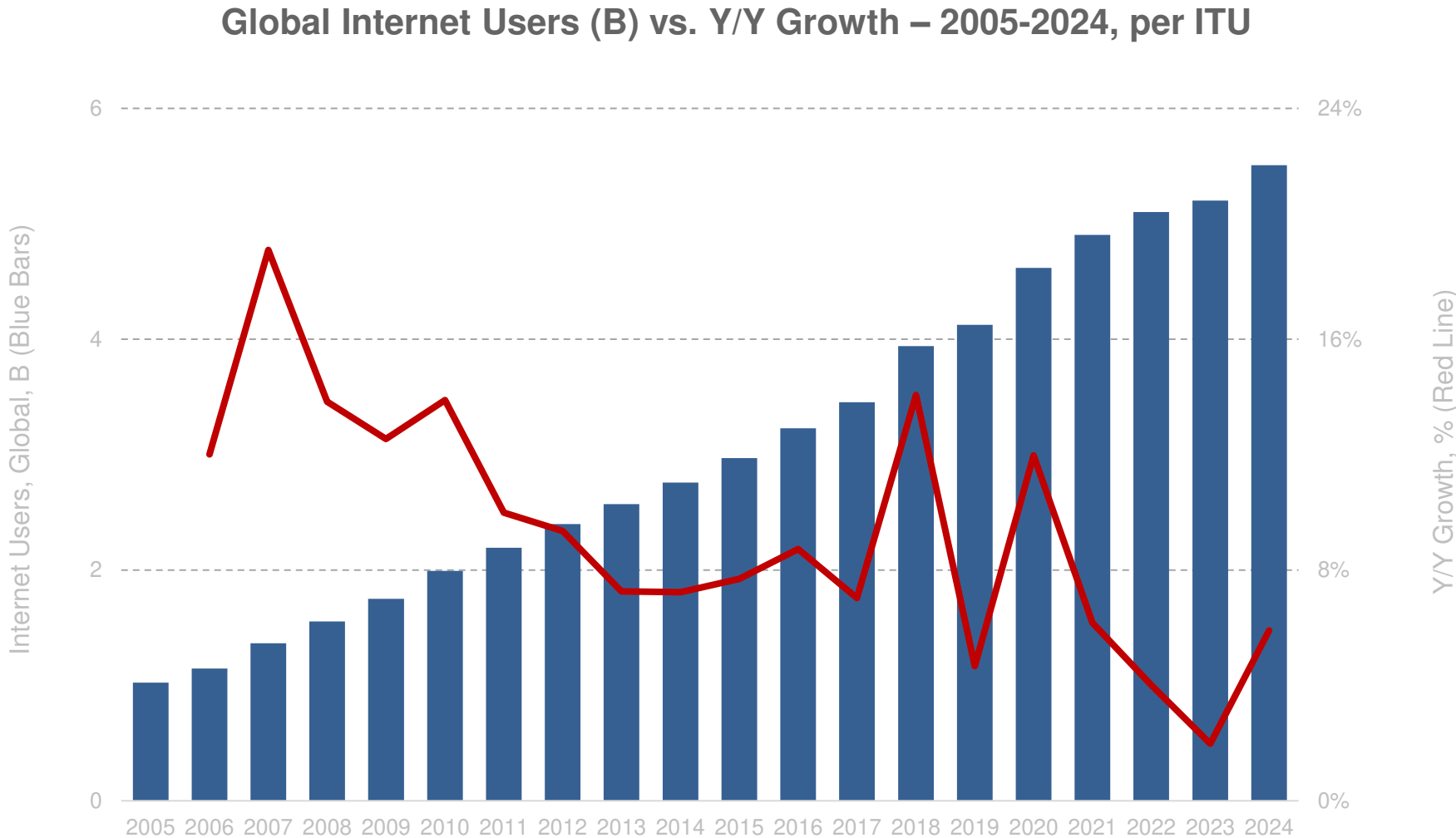
Global Internet Penetration by Population Density =  
83% of Urban Dwellers Online vs. 48% Rural

按人口密度划分的全球互联网普及率 =83% 的城市居民在  
线 vs. 48% 的农村居民

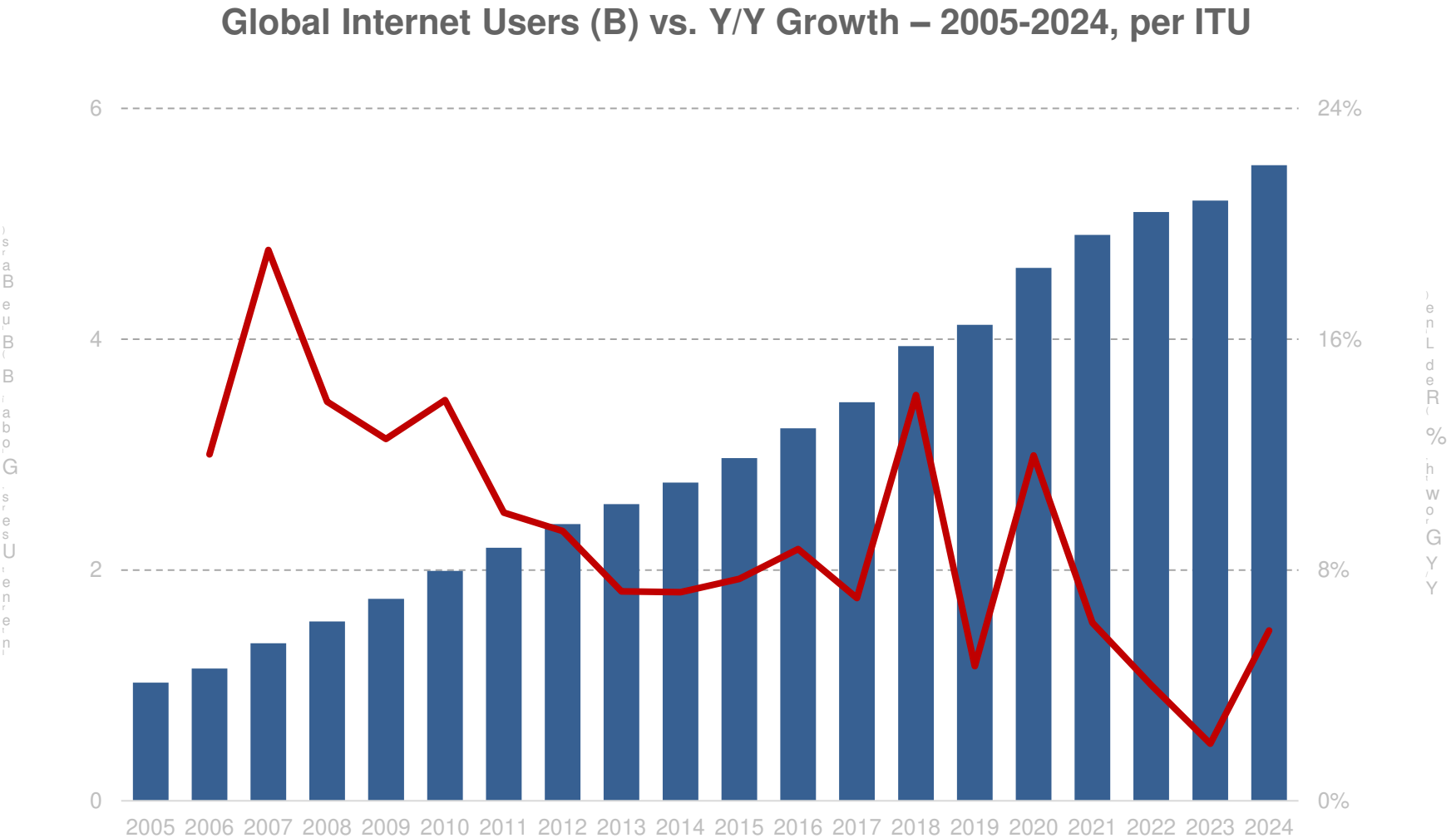


Global Internet Users @ 5.5B =  
+6% Y/Y & Accelerating, per ITU

Global Internet Users @ 5.5B =  
+6% Y/Y & Accelerating, per ITU



Source: United Nations / International Telecommunications Union (3/25)

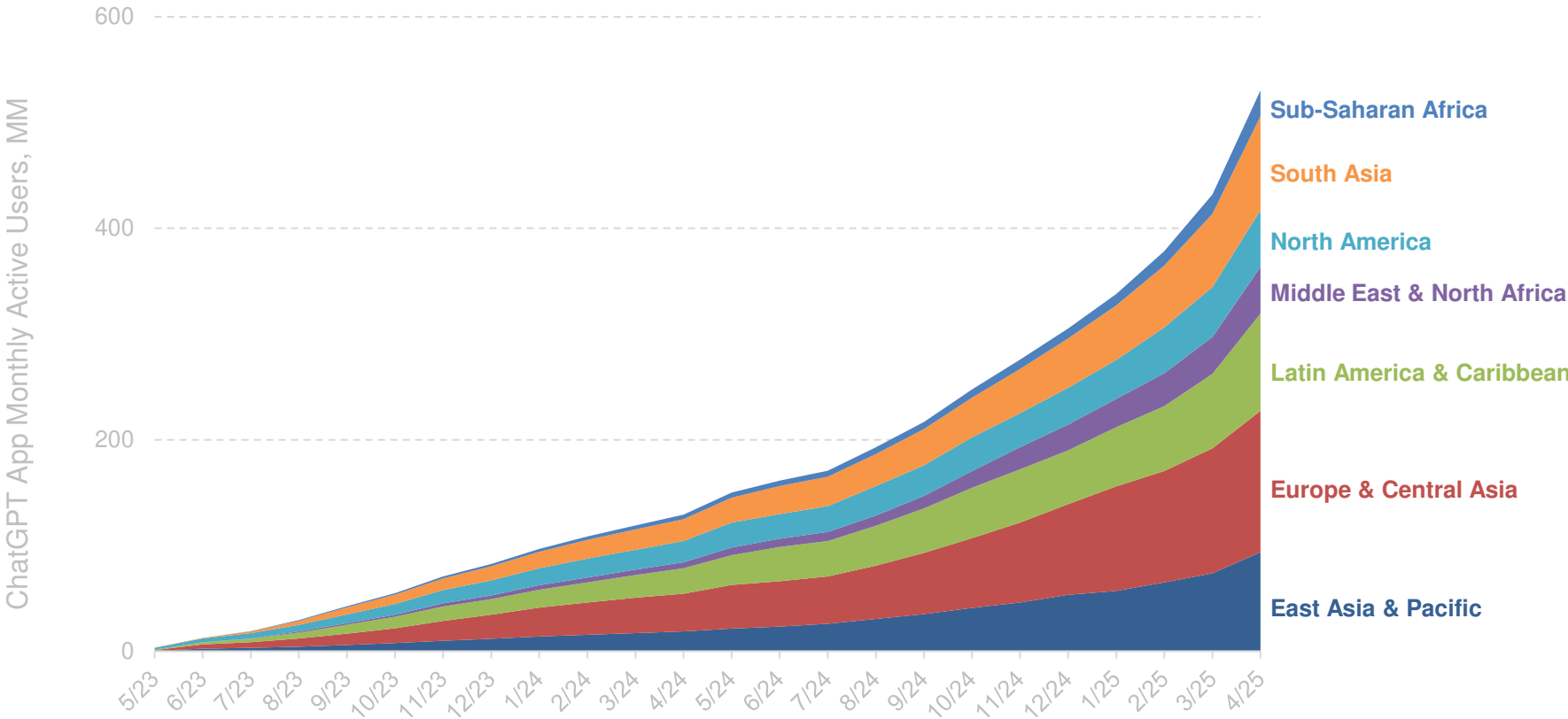


来源：联合国 / 国际电信联盟（3/25）

# ChatGPT Mobile App @ 530MM MAUs in Twenty-Three Months = Global Growth We Have Not Seen Likes Of Before

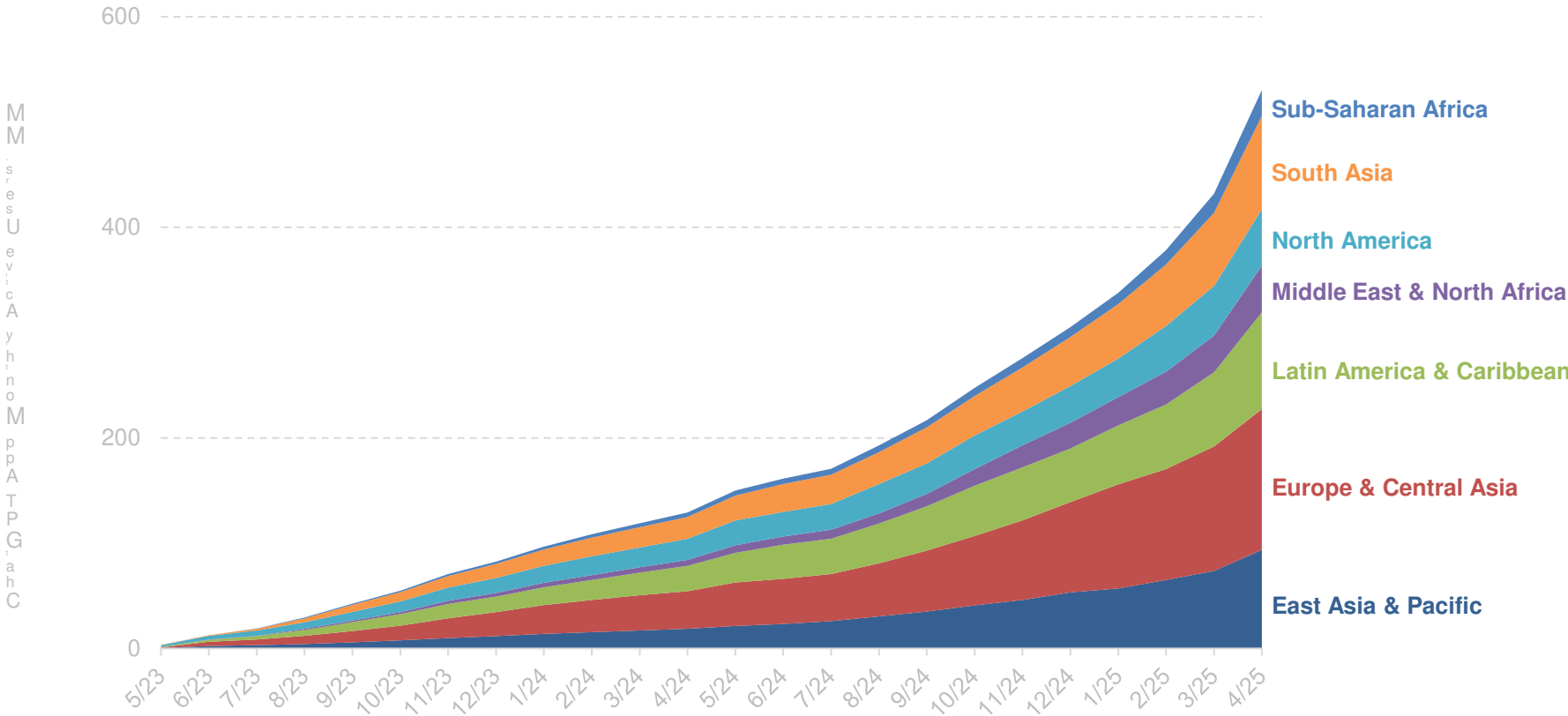
# ChatGPT 移动应用 @ 23 个月内达到 5.3 亿 MAU = 我们从未见过的全球增长速度

ChatGPT App Monthly Active Users (MAUs) (MM) – 5/23-4/25, per Sensor Tower



Note: Regions are per United Nations definitions. ChatGPT app not available in China, Russia and select other countries as of 5/25. Includes only Android, iPhone & iPad users. Figures may understate true ChatGPT user base (e.g., desktop or mobile webpage users). Data for standalone app only. Source: Sensor Tower (5/25)

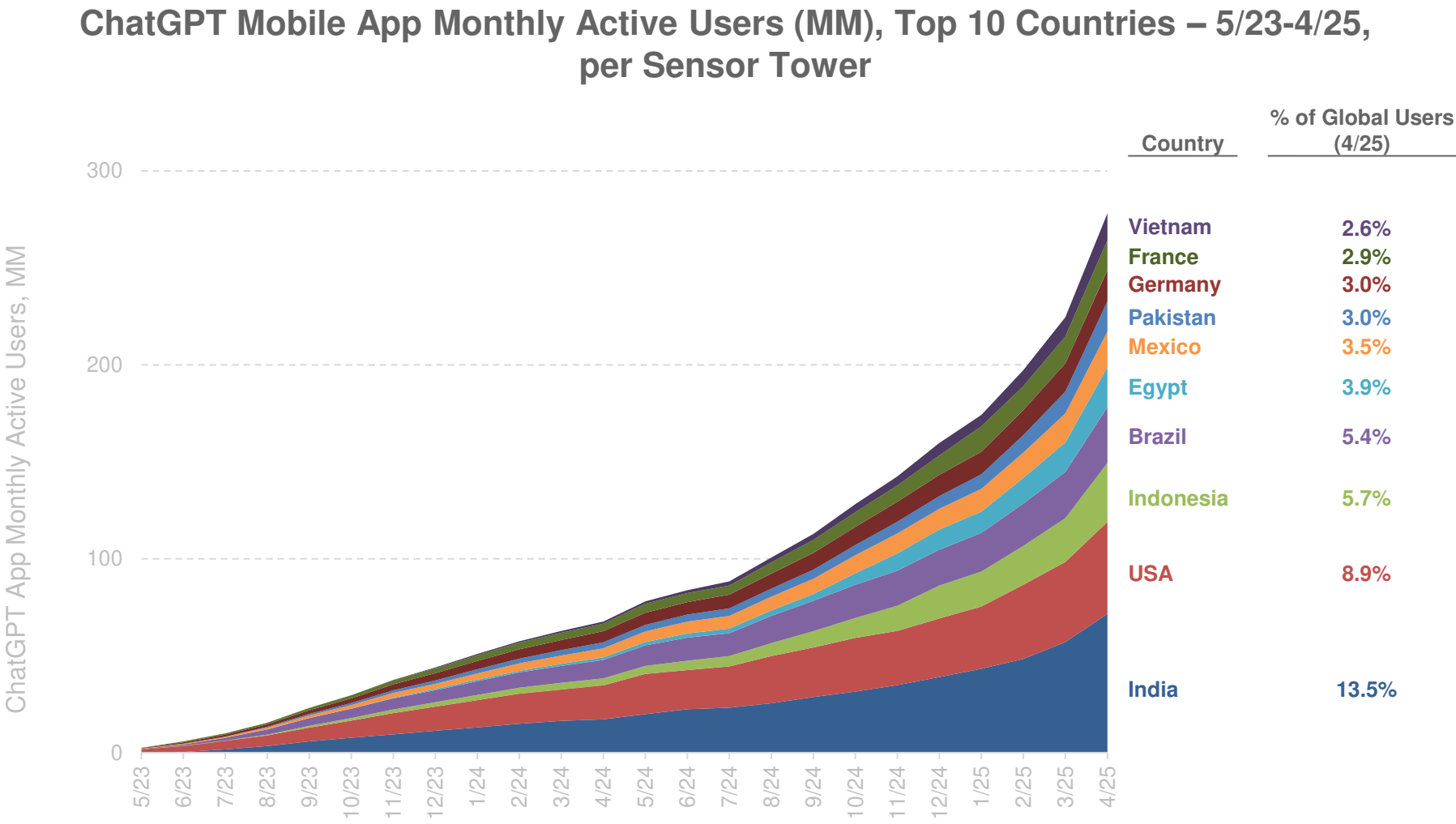
ChatGPT App Monthly Active Users (MAUs) (MM) – 5/23-4/25, per Sensor Tower



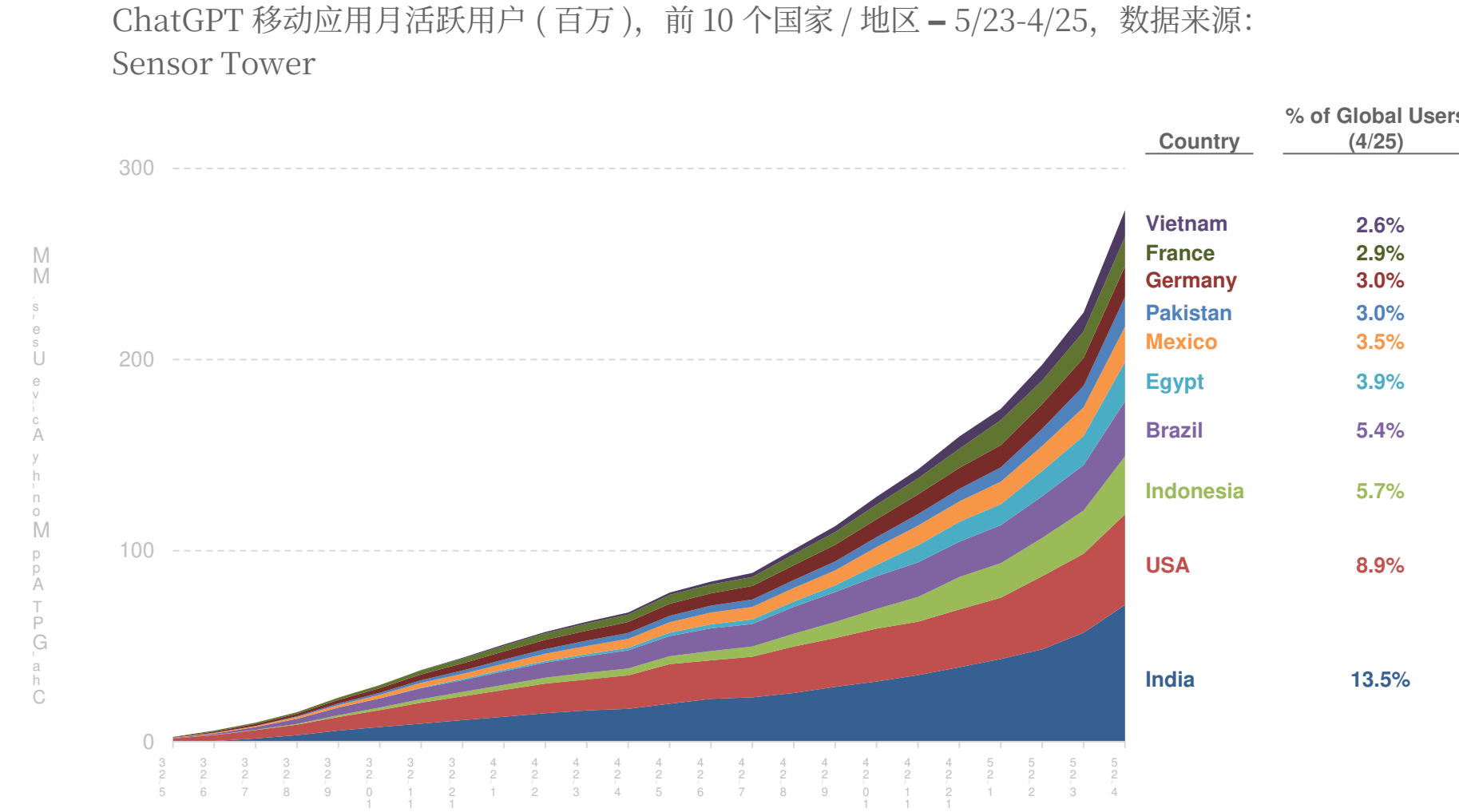
注意：区域划分依据联合国定义。截至 5 月 25 日，ChatGPT 应用程序在中国、俄罗斯和部分其他国家 / 地区不可用。仅包含 Android、iPhone 和 iPad 用户。数据可能低估了真实的 ChatGPT 用户群（例如，桌面或移动网页用户）。数据仅适用于独立应用程序。来源：Sensor Tower（5/25）

ChatGPT Mobile App – Top User Countries = India @ 14%...USA @ 9%...Indonesia @ 6%, per Sensor Tower

ChatGPT 移动应用 – 热门用户国家 / 地区 = 印度 @ 14%... 美国 @ 9%... 印度尼西亚 @ 6%，数据来源：Sensor Tower



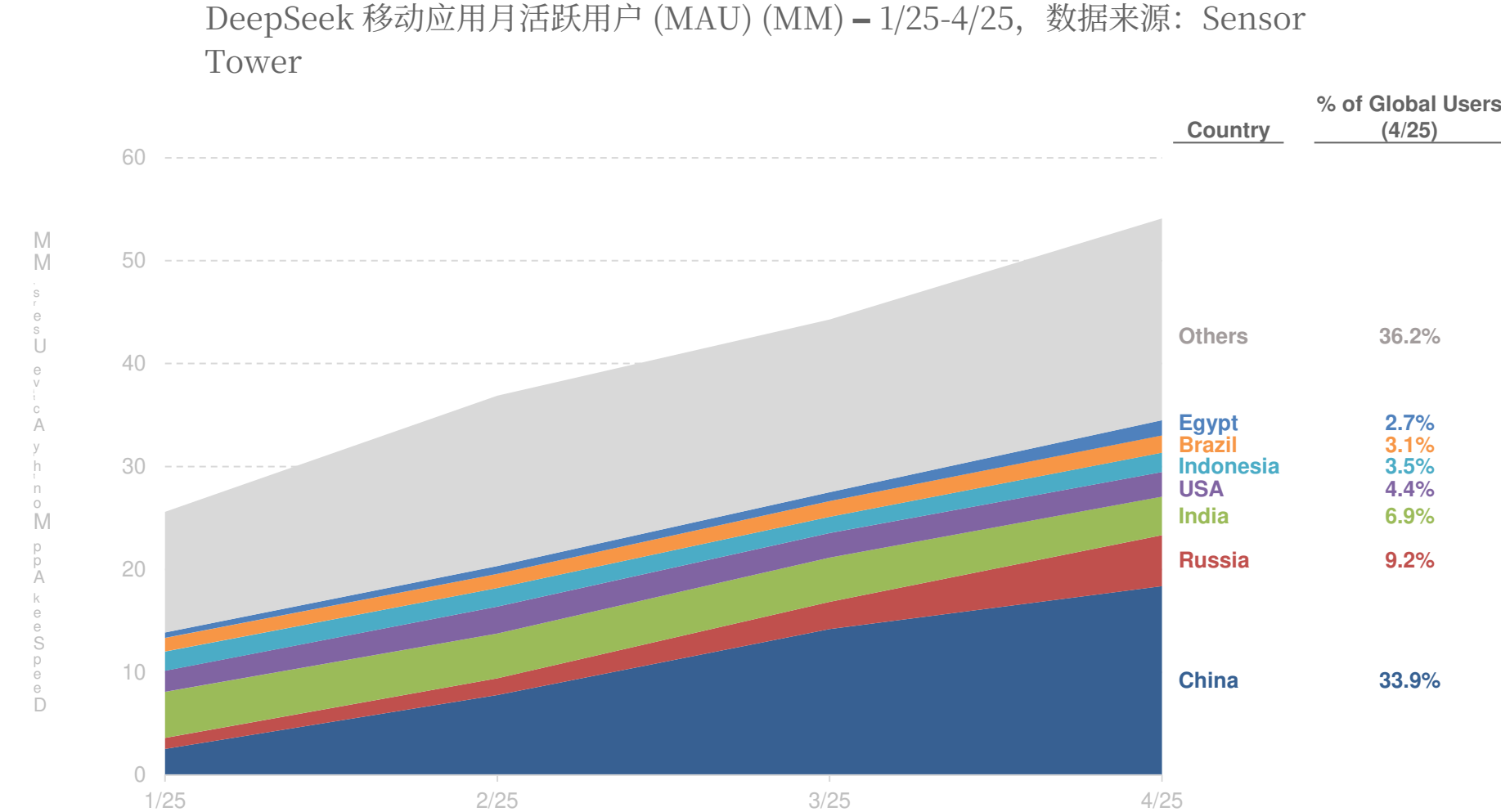
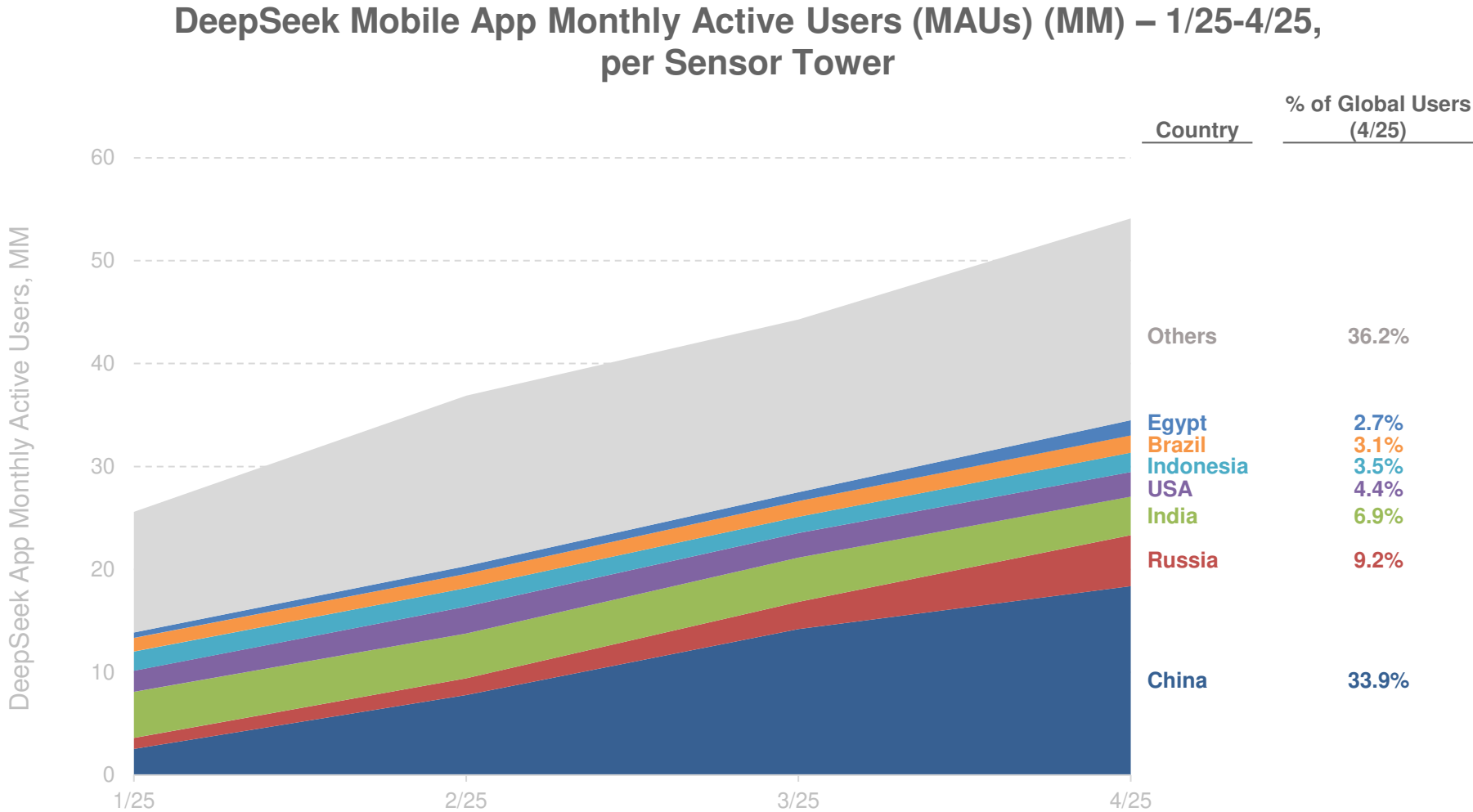
Note: Regions are per United Nations definitions. ChatGPT app not available in China, Russia and select other countries as of 5/25. Includes only Android, iPhone & iPad users. Figures may understate true ChatGPT user base (e.g., desktop or mobile webpage users). Data for standalone app only. Source: Sensor Tower (5/6/25)



注意：地区划分遵循联合国定义。截至 5 月 25 日，ChatGPT 应用程序在中国、俄罗斯和部分其他国家 / 地区不可用。仅包括 Android、iPhone 和 iPad 用户。数据可能低估了 ChatGPT 的真实用户群（例如，桌面或移动网页用户）。数据仅适用于独立应用程序。来源：Sensor Tower (5/6/25)

# DeepSeek Mobile App @ 54MM MAUs in Four Months = Growth Concentrated in China (34% Users) & Russia (9%)

# DeepSeek 移动应用 @ 54MM MAU 四个月内 = 增长集中在中国（34% 用户）和俄罗斯（9%）



Note: Regions are per United Nations definitions. Includes only Android, iPhone & iPad users. Figures may understate true DeepSeek user base (e.g., desktop or mobile webpage users). Data for standalone app only. Data may be incomplete for China, Russia, and select other countries due to informational restrictions. Source: Sensor Tower (5/6/25)

注意：区域划分按照联合国定义。仅包括 Android、iPhone 和 iPad 用户。数据可能低估了 DeepSeek 的真实用户群（例如，桌面或移动网页用户）。数据仅适用于独立应用。由于信息限制，中国、俄罗斯和部分其他国家 / 地区的数据可能不完整。来源：Sensor Tower (5/6/25)

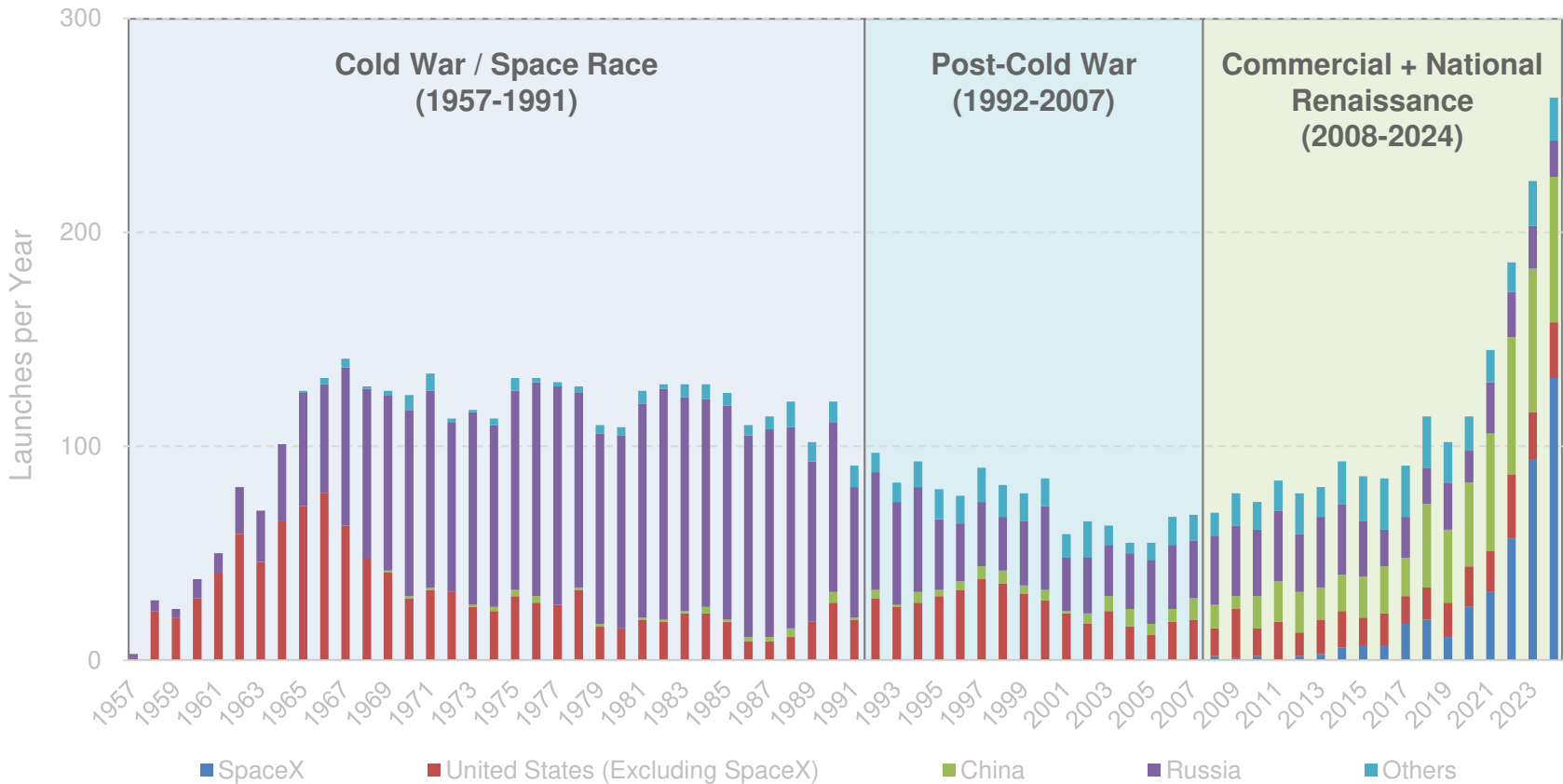
New Internet User Growth =  
Enabled by AI + Satellites

新增互联网用户增长 =  
由 AI + 卫星驱动

# Orbital / Satellite Launch Market Share, Global = SpaceX Rising

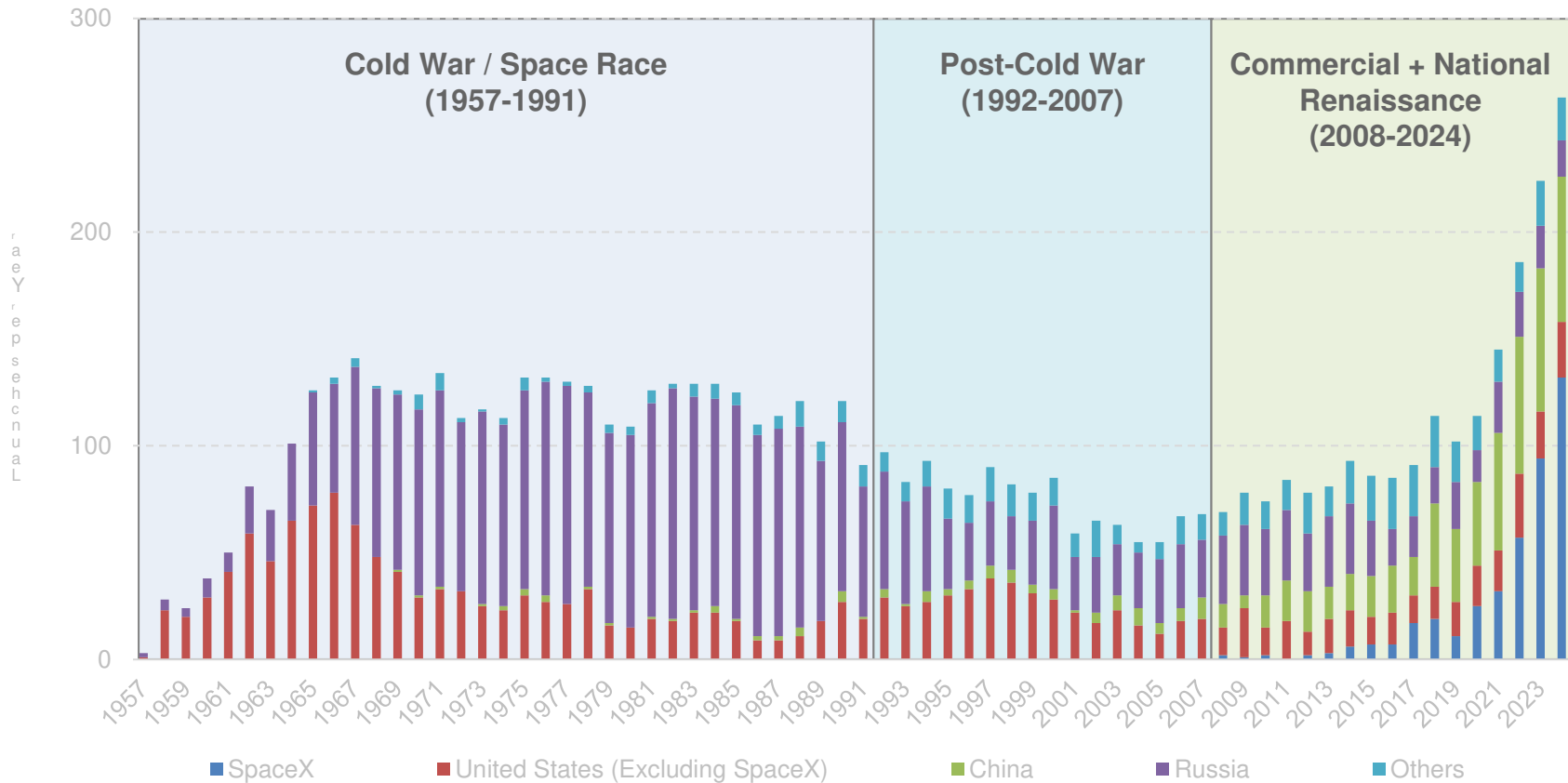
# 全球轨道 / 卫星发射市场份额 =SpaceX 崛起

Orbital Launches by Year & Country – 1957-2025, per SpaceX, Space Stats & USA FAA



Note: Orbital launches from other celestial bodies than Earth are not included (e.g., Apollo LM ascents from the Moon's surface).  
Source: SpaceX public announcements (1/25), Space Stats (3/25), USA Federal Aviation Administration (3/25)

按年份和国家 / 地区划分的轨道发射次数 – 1957-2025，数据来源：SpaceX、Space Stats 和 USA FAA

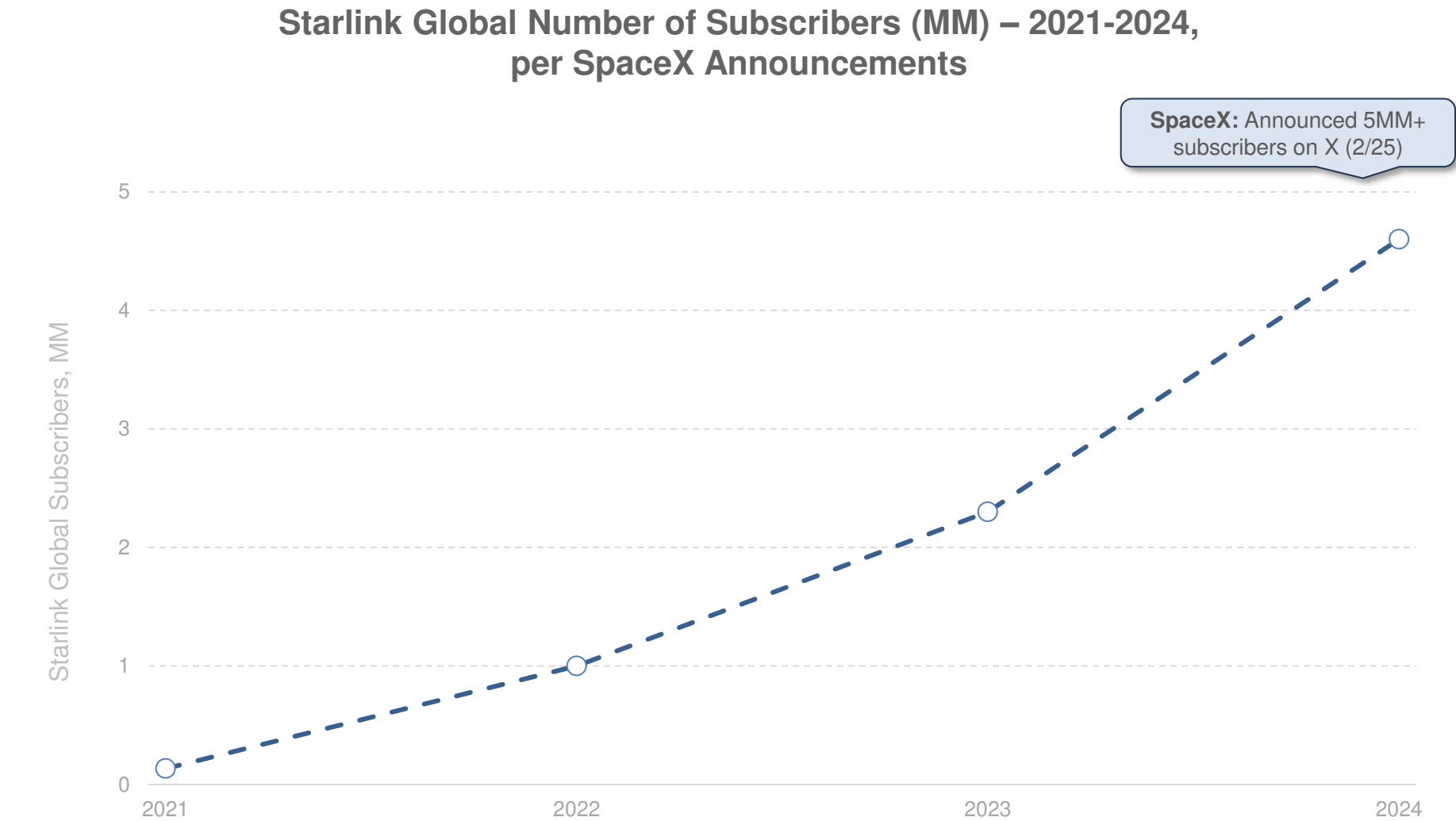


Note: Orbital launches from other celestial bodies than Earth are not included (e.g., Apollo LM ascents from the Moon's surface).  
Source: SpaceX public announcements (1/25), Space Stats (3/25), USA Federal Aviation Administration (3/25)

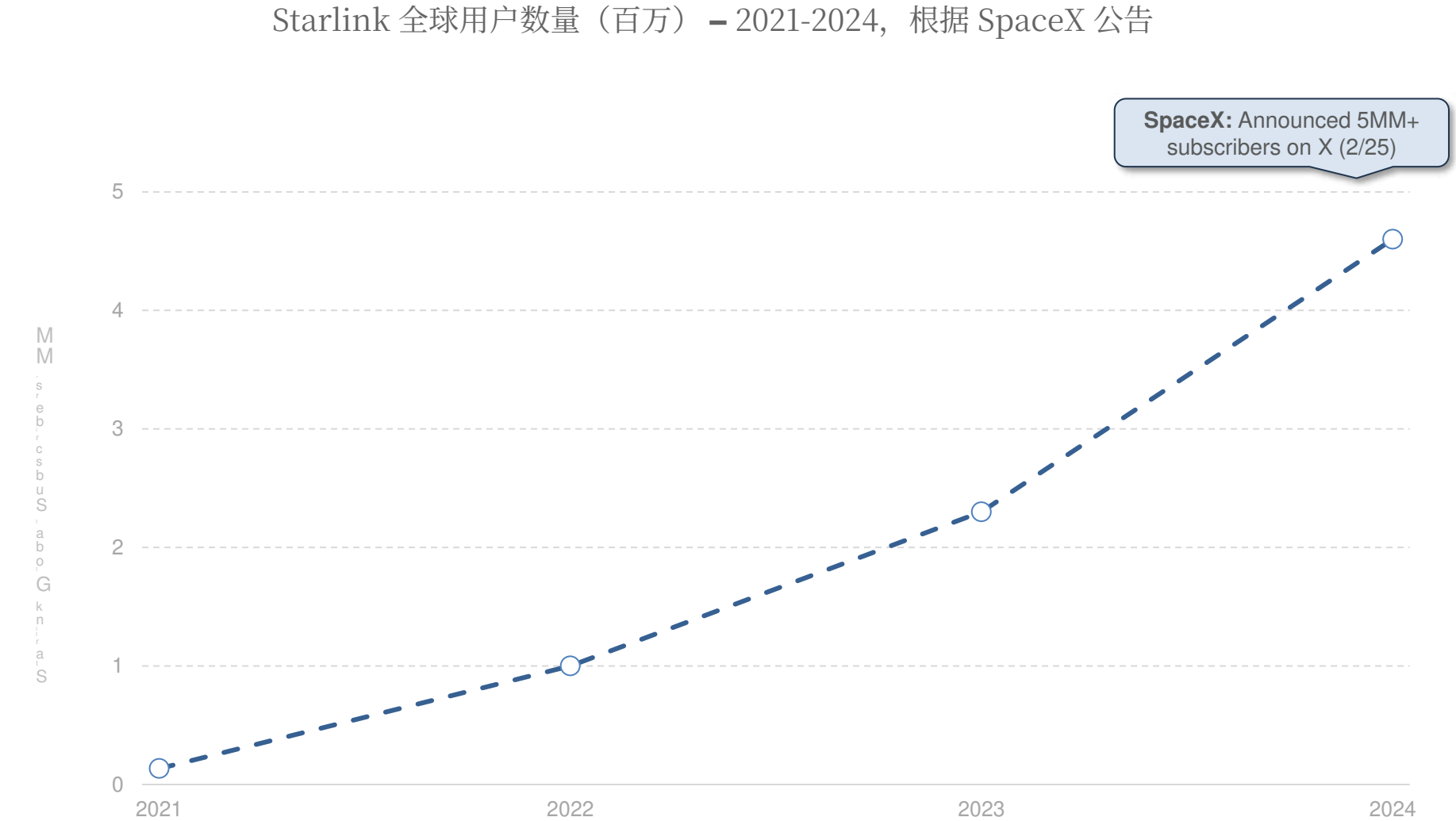


SpaceX Starlink @ 5MM+ Subscribers =  
+202% Annual Growth Over 3.2 Years

SpaceX Starlink @ 5MM+ 订阅者 =+202  
3.2 年内年增长率



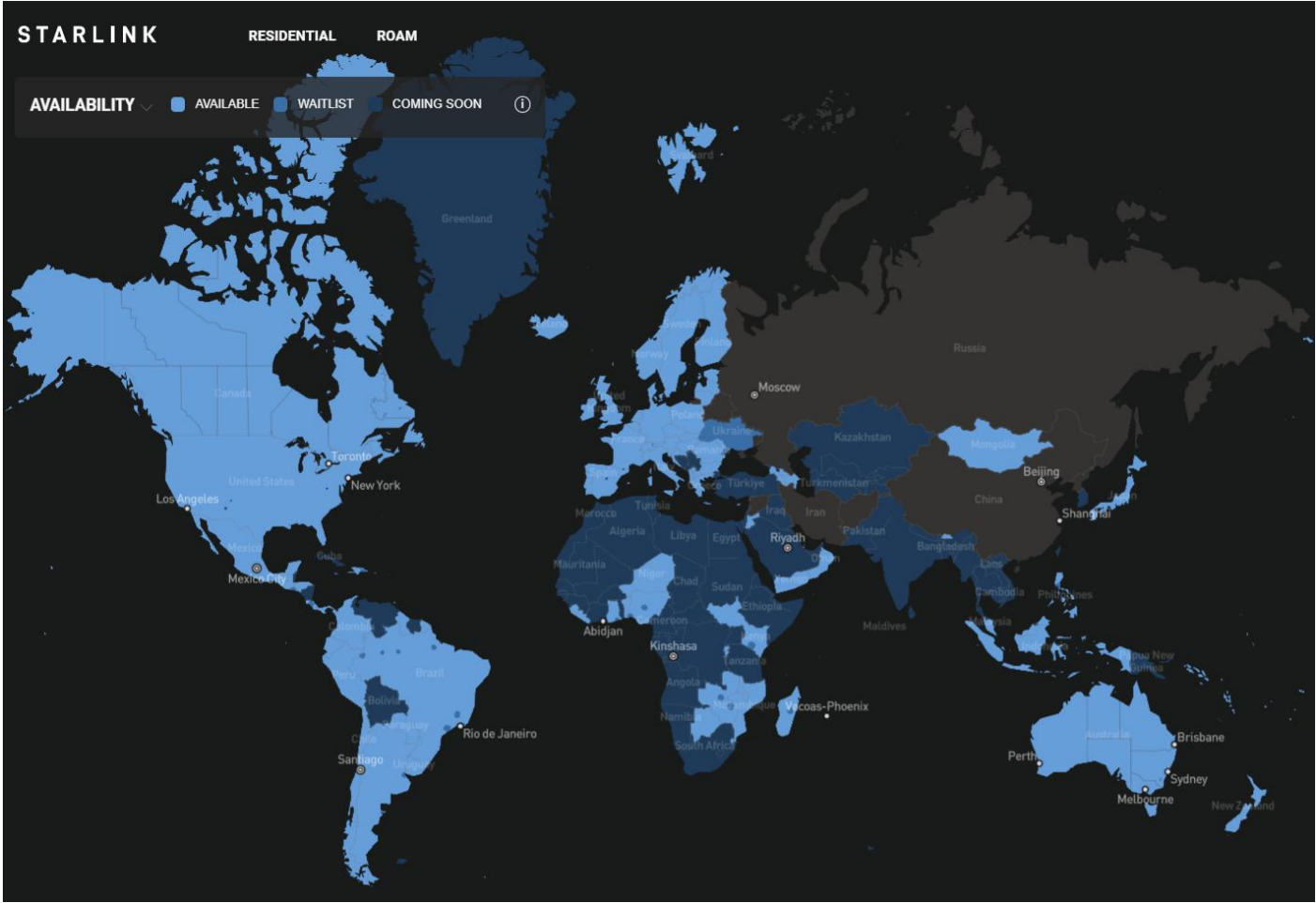
Source: SpaceX public announcements



Source: SpaceX public announcements

# SpaceX Starlink Ecosystem = Coverage Expanding Globally

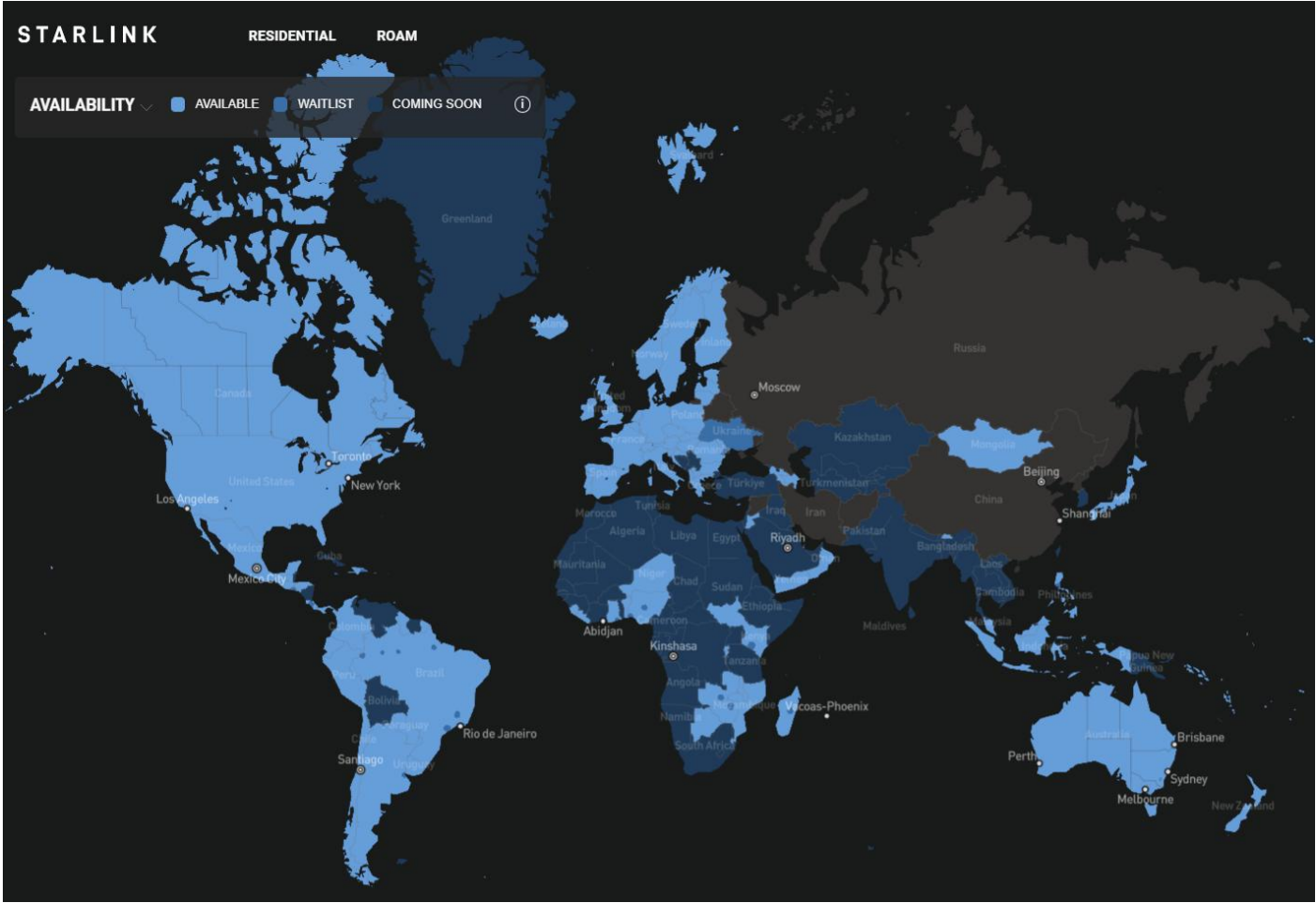
Starlink Global Coverage – 5/25, per SpaceX



Source: SpaceX website (5/25)

# SpaceX Starlink 生态系统 = 覆盖范围全球扩展

Starlink 全球覆盖范围 – 5/25, 来自 SpaceX



Source: SpaceX website (5/25)

# Starlink = Unlocking Previously-Inaccessible Internet Access in AI Era

## Select Global Starlink Use Cases – 4/25, per SpaceX



### Coco, Monterrey, Mexico

*Starlink's technology has enabled Coco's operations, delivering high-speed, reliable internet that bridges the digital divide in rural Mexico. Through our streamlined community WiFi services, we're not just offering connectivity, we're opening a window to the world for hundreds in remote areas. With Starlink, we've boosted connection speeds and efficiency, transforming disconnected regions into digitally engaged communities.*



### Chile School District

*[Our] school went from slow, ineffective connectivity for even 2-3 computer stations, to having high-speed internet where all 36 of our children can have effective internet connectivity simultaneously...a class-changing event for our teachers and students.*



### Brightline Trains, USA

*Starlink gave us the new beginning we were looking for. It gave us connectivity we can be proud to share with our guests. It gave us the knowledge we needed to continue to build better train connectivity beyond the satellite [internet] itself...and, most of all, it gave us a new beginning for train enthusiasts to get excited about because it is doable, it is maintainable, [and] it is as exciting as it seems.*



### Seaspan Corporation, Global

*Deploying SpaceX Starlink's low Earth orbit, low-latency, high bandwidth service across our fleet is a major milestone in addressing connectivity challenges in an industry with a global and mobile workforce. It allows us to treat our vessels no differently than remote offices, supporting crew safety and wellness – and it enables us to develop new solutions that were technically and financially unviable just a few years ago.*

# Starlink = Unlocking Previously-Inaccessible Internet Access in AI Era

## 选择全球 Starlink 使用案例 – 4/25, 根据 SpaceX



### Coco, Monterrey, Mexico

*Starlink's technology has enabled Coco's operations, delivering high-speed, reliable internet that bridges the digital divide in rural Mexico. Through our streamlined community WiFi services, we're not just offering connectivity, we're opening a window to the world for hundreds in remote areas. With Starlink, we've boosted connection speeds and efficiency, transforming disconnected regions into digitally engaged communities.*



### Chile School District

*[Our] school went from slow, ineffective connectivity for even 2-3 computer stations, to having high-speed internet where all 36 of our children can have effective internet connectivity simultaneously...a class-changing event for our teachers and students.*



### Brightline Trains, USA

*Starlink gave us the new beginning we were looking for. It gave us connectivity we can be proud to share with our guests. It gave us the knowledge we needed to continue to build better train connectivity beyond the satellite [internet] itself...and, most of all, it gave us a new beginning for train enthusiasts to get excited about because it is doable, it is maintainable, [and] it is as exciting as it seems.*



### Seaspan Corporation, Global

*Deploying SpaceX Starlink's low Earth orbit, low-latency, high bandwidth service across our fleet is a major milestone in addressing connectivity challenges in an industry with a global and mobile workforce. It allows us to treat our vessels no differently than remote offices, supporting crew safety and wellness – and it enables us to develop new solutions that were technically and financially unviable just a few years ago.*

- 1 Seem Like Change Happening Faster Than Ever?  
Yes, It Is
- 2 AI User + Usage + CapEx Growth =  
Unprecedented
- 3 AI Model Compute Costs High / Rising + Inference Costs Per Token Falling =  
Performance Converging + Developer Usage Rising
- 4 AI Usage + Cost + Loss Growth =  
Unprecedented
- 5 AI Monetization Threats =  
Rising Competition + Open-Source Momentum + China's Rise
- 6 AI & Physical World Ramps =  
Fast + Data-Driven
- 7 Global Internet User Ramps Powered by AI from Get-Go =  
Growth We Have Not Seen Likes of Before
- 8 AI & Work Evolution =  
Real + Rapid

- 1 变化似乎比以往任何时候都快？是的，确实如此
- 2 AI 用户 + 使用量 + 资本支出增长 = 前所未有
- 3 AI 模型计算成本高 / 上升 + 每次 Token 的推理成本下降 = 性能趋同 + 开发者使用量上升
- 4 AI 使用量 + 成本 + 亏损增长 = 前所未有
- 5 人工智能货币化威胁 = 日益激烈的竞争 + 开源势头 + 中国的崛起
- 6 AI 与物理世界的斜坡 = 快 + 数据驱动
- 7 Global Internet User Ramps Powered by AI from Get-Go = 我们以前从未见过的增长
- 8 AI 与工作变革 = 真实 + 快速

AI is foundationally changing the way we work. Alongside growth in physical automation (think adoption of robots and drones), we are now also seeing the rise of cognitive automation, where AI systems can reason, create, and solve problems. The ramifications are widespread.

The pace of improvement in AI's cognitive ability is astounding. In the three years since ChatGPT’s 11/22 public launch, we've gone from the reasoning capabilities of a high school student to those of a PhD candidate. Professions centered on intaking large bodies of structured, historical data and outputting rules-based decisions and judgement, fall squarely in the core competency of generative AI. In this emerging landscape, a unit of labor could shift from human hours to computational power. Data centers and foundation models – in many instances – could dictate the availability and quality of certain types of labor.

As a result, some tout an 'agentic future' where AI agents replace humans in many white-collar jobs. Although possible, history and pattern recognition suggest the role of humans is enduring and compelling. Technology-forward leaps have typically driven productivity and efficiency gains and more – but new – jobs. That said, this time it’s happening faster.

In an extreme, entirely agentic future, humans maintain a role in the system, pivoting towards oversight, guidance, and training. Imagine facilities filled with humans teaching robots intricate movements or offices full of workers providing reinforcement learning\* human feedback (RLHF) to optimize algorithms. This is not conjecture. Companies like Physical Intelligence and Scale AI, respectively, are building powerful businesses based on this view of the world.

The idea of the human workforce re-configured to teach and refine machines as a primary function might sound dystopic. But it’s worth remembering historical parallels. Fifty years ago, this prospect of rows of cubicles and uniformed office workers sitting quietly in front of LED computers ten hours a day likely sounded equally dystopic. Yet here we are.

Technology has constantly redefined and evolved the nature of work and productivity...AI is no different.

*\*Reinforcement Learning = An ML approach where agents learn by receiving rewards or penalties for actions.*

人工智能正在从根本上改变我们的工作方式。随着物理自动化（如机器人和无人机的应用）的增长，我们现在也看到了认知自动化的兴起，人工智能系统可以进行推理、创造和解决问题。其影响是广泛的。

人工智能认知能力提升的速度令人震惊。自从 ChatGPT 于 22 年 11 月公开发布以来的三年里，我们已经从高中生的推理能力发展到了博士候选人的水平。以摄取大量结构化历史数据并输出基于规则的决策和判断为中心的职业，完全属于生成式人工智能的核心能力。在这种新兴的格局下，一个劳动单位可能会从人工时转变为计算能力。数据中心和基础模型 – 在许多情况下 – 可能会决定某些类型劳动力的可用性和质量。

因此，有些人鼓吹 “ 代理未来 ”，即人工智能代理将在许多白领工作中取代人类。虽然有可能，但历史和模式识别表明，人类的角色是持久且引人注目的。技术上的飞跃通常会推动生产力和效率的提高，并创造更多 – 但新的 – 工作。也就是说，这次发生的速度更快。

在一个极端的、完全自主的未来，人类在系统中保持着一定的角色，转向监督、指导和培训。想象一下，在一些设施里，人类教机器人复杂的动作；或者在一些办公室里，员工提供强化学习 \* 人类反馈 (RLHF) 来优化算法。这并非臆测。Physical Intelligence 和 Scale AI 等公司分别正在基于这种世界观建立强大的业务。

将人类劳动力重新配置为以教导和改进机器作为主要职能的想法可能听起来像反乌托邦。但值得回顾一下历史上的相似之处。五十年前，一排排小隔间和穿着制服的办公室工作人员每天安静地坐在 LED 电脑前十个小时的前景可能听起来同样像反乌托邦。但我们现在就在这里。

技术不断地重新定义和发展着工作和生产力的本质 …… 人工智能也不例外。

*\*Reinforcement Learning = 一种 ML 方法，在这种方法中，智能体通过接收行动的奖励或惩罚来进行学习。*



# AI Impact on Business = Diverse & Broad

# AI 对业务的影响 = 多样化且广泛

Industries That Could Be Affected by AI, per NVIDIA



Note: Global data shown. Source: NVIDIA

NVIDIA 认为可能受到 AI 影响的行业



注意：显示的是全球数据。来源：NVIDIA

# AI In Workforce – Shopify = *Reflexive AI Usage Is Now a Baseline Expectation...*



*We are entering a time where more merchants and entrepreneurs could be created than any other in history.  
We often talk about bringing down the complexity curve to allow more people to choose this as a career.  
Each step along the entrepreneurial path is rife with decisions requiring skill, judgement and knowledge.  
Having AI alongside the journey and increasingly doing not just the consultation,  
but also doing the work for our merchants is a mind-blowing step function change here.*

*Our task here at Shopify is to make our software unquestionably the best canvas on which to develop  
the best businesses of the future. We do this by keeping everyone cutting edge and bringing all the best tools  
to bear so our merchants can be more successful than they themselves used to imagine.  
For that we need to be absolutely ahead.*

*Reflexive AI usage is now a baseline expectation at Shopify.*

*Maybe you are already there and find this memo puzzling. In that case you already use AI as a thought partner,  
deep researcher, critic, tutor, or pair programmer. I use it all the time, but even I feel I'm only scratching the surface.  
It's the most rapid shift to how work is done that I've seen in my career...*

*...Using AI effectively is now a fundamental expectation of everyone at Shopify.  
It's a tool of all trades today, and will only grow in importance. Frankly, I don't think it's feasible to opt out of learning the skill of  
applying AI in your craft; you are welcome to try, but I want to be honest I cannot see this working out today,  
and definitely not tomorrow.*

*Stagnation is almost certain, and stagnation is slow-motion failure. If you're not climbing, you're sliding...*

Shopify Co-Founder & CEO Tobias Lütke in Internal Memo on AI – 3/25

Source: Tobi Lutke via X (4/25), Shopify

# AI In Workforce – Shopify = 反射性人工智能的使用现在已 成为基本期望 .....



我们正在进入一个时代，在这个时代里，可以创造出比历史上任何时候都多的商家和企业家。我们经常谈论降低复杂性曲线，以便让更多的人选择将其作为职业。创业道路上的每一步都充满了需要技能、判断和知识的决策。在旅程中拥有人工智能，并且越来越多地不仅仅是咨询，而且还为我们的商家做工作，这是一个令人难以置信的阶跃函数变化。

我们在 Shopify 的任务是让我们的软件无疑是开发未来最佳业务的最佳平台。我们通过让每个人都保持领先地位，并提供所有最好的工具来做到这一点，以便我们的商家能够比他们自己想象的更成功。为此，我们需要绝对领先。

在 Shopify，反射性 AI 的使用现在已成为一项基本要求。

也许你已经身处其中，并且觉得这份备忘录令人困惑。如果是这样，你已经将 AI 用作思考伙伴、深度研究员、评论家、导师或结对程序员。我一直都在使用它，但即使是我也觉得我只是触及了表面。这是我职业生涯中见过的最快速的工作方式转变 .....

..... 现在，有效使用 AI 是 Shopify 每位员工的一项基本要求。它现在是一种万能工具，而且只会变得越来越重要。坦率地说，我认为选择不学习在你自己的专业中应用 AI 的技能是不可行的；欢迎你尝试，但我想坦诚地告诉你，我认为这在今天行不通，而且肯定在明天也行不通。停滞几乎是必然的，而停滞是慢动作失败。如果你不往上爬，你就在下滑 .....

Shopify 联合创始人兼首席执行官 Tobias Lütke 在关于 AI 的内部备忘录中 – 3/25

来源: Tobi Lutke via X (4/25), Shopify



# ...AI In Workforce – Duolingo = *Duolingo Is Going to be AI-First*



*I've said this in Q&As and many meetings, but I want to make it official: Duolingo is going to be AI-first.*

*AI is already changing how work gets done. It's not a question of if or when. It's happening now. When there's a shift this big, the worst thing you can do is wait. In 2012, we bet on mobile. While others were focused on mobile companion apps for websites, we decided to build mobile-first because we saw it was the future. That decision helped us win the 2013 iPhone App of the Year and unlocked the organic word-of-mouth growth that followed...*

*...AI isn't just a productivity boost. It helps us get closer to our mission. To teach well, we need to create a massive amount of content, and doing that manually doesn't scale. One of the best decisions we made recently was replacing a slow, manual content creation process with one powered by AI. Without AI, it would take us decades to scale our content to more learners. We owe it to our learners to get them this content ASAP...*

*...Being AI-first means we will need to rethink much of how we work. Making minor tweaks to systems designed for humans won't get us there...We can't wait until the technology is 100% perfect. We'd rather move with urgency and take occasional small hits on quality than move slowly and miss the moment.*

*We'll be rolling out a few constructive constraints to help guide this shift...:*

- *...AI use will be part of what we look for in hiring*
- *AI use will be part of what we evaluate in performance reviews*
- *Headcount will only be given if a team cannot automate more of their work*
- *Most functions will have specific initiatives to fundamentally change how they work...*

Duolingo Co-Founder & CEO Luis von Ahn in All-Hands Memo on AI – 4/25

Source: Duolingo via LinkedIn (4/25)

# ...AI In Workforce – Duolingo = *Duolingo Is Going to be AI-First*



我在问答环节和多次会议中都说过这一点，但我想正式宣布：Duolingo 将会是 AI-first。

AI 已经在改变工作完成的方式。这已经不是一个是否或者何时发生的问题了。它正在发生。当出现如此巨大的转变时，最糟糕的事情就是等待。2012 年，我们押注于移动设备。当其他人专注于网站的移动配套应用时，我们决定首先构建移动设备，因为我们看到了它的未来。这个决定帮助我们赢得了 2013 年 iPhone 年度应用奖，并开启了随之而来的口碑式的自然增长 ……*mouth growth that followed...*

…AI 不仅仅能提高生产力。它帮助我们更接近我们的使命。为了教好，我们需要创建大量的内容，而手动完成这项工作是无法扩展的。我们最近做出的最好的决定之一就是用 AI 驱动的内容创作流程取代了缓慢的手动内容创作流程。如果没有 AI，我们需要几十年才能将我们的内容扩展到更多的学习者。为了尽快给学习者提供这些内容，我们有责任这样做 ……

*...Being AI-first* 意味着我们需要重新思考我们的工作方式。对为人类设计的系统进行小的调整无法实现这一目标 …

… 我们不能等到技术达到 100% 完美。我们宁愿紧急行动，偶尔在质量上受到小的影响，也不愿行动缓慢而错过时机。

我们将推出一些建设性的约束，以帮助指导这种转变 ……:

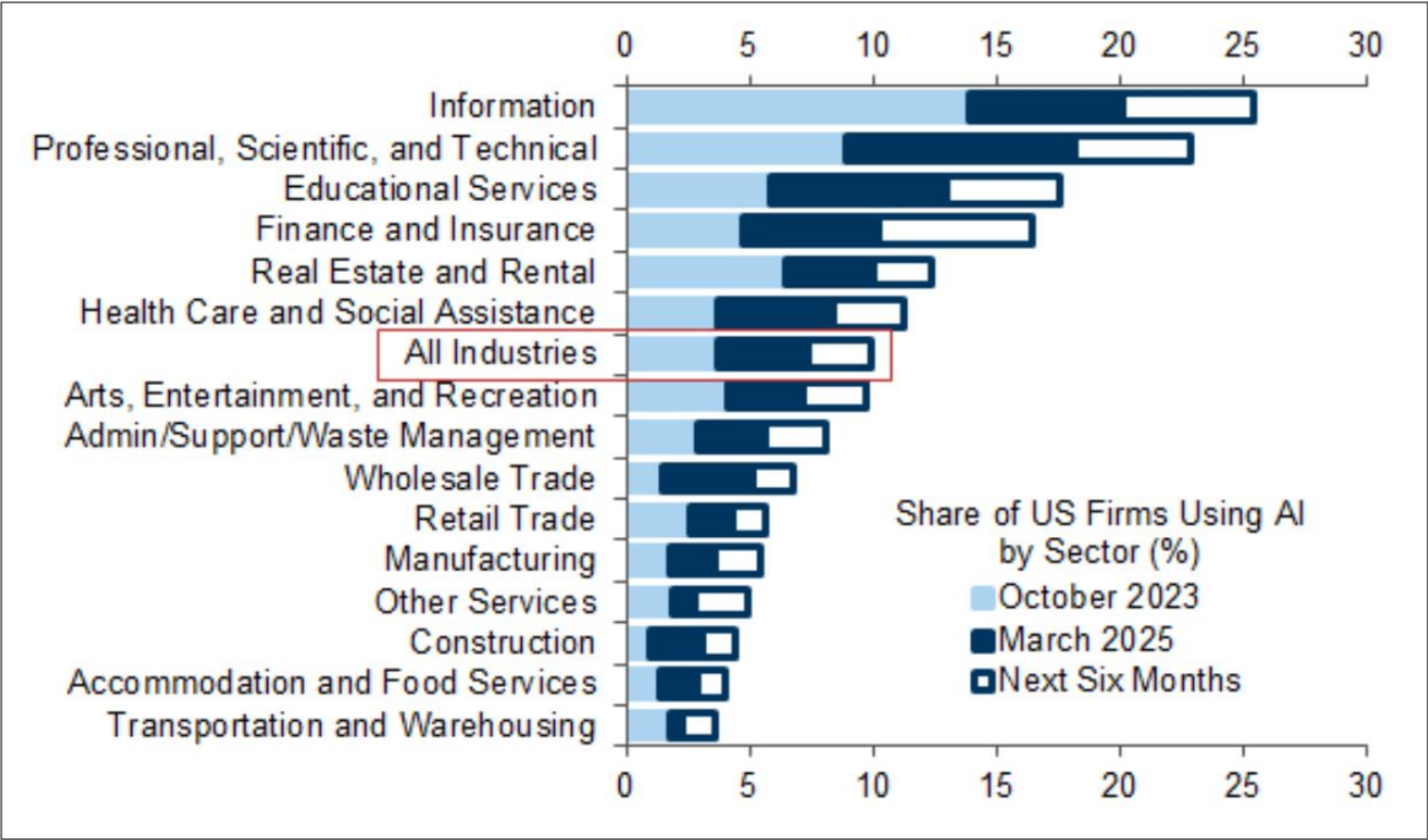
- ……AI 的使用将成为我们招聘时考察的一部分
- AI 的使用将成为我们在绩效评估中评估的一部分
- 只有当团队无法自动化更多工作时，才会提供人员编制
- 大多数职能部门将采取具体举措，从根本上改变其工作方式 ……

Duolingo 联合创始人兼 CEO Luis von Ahn 在全体员工 AI 备忘录中 – 4/25

Source: Duolingo via LinkedIn (4/25)

# AI Adoption @ USA Firms = Rising...

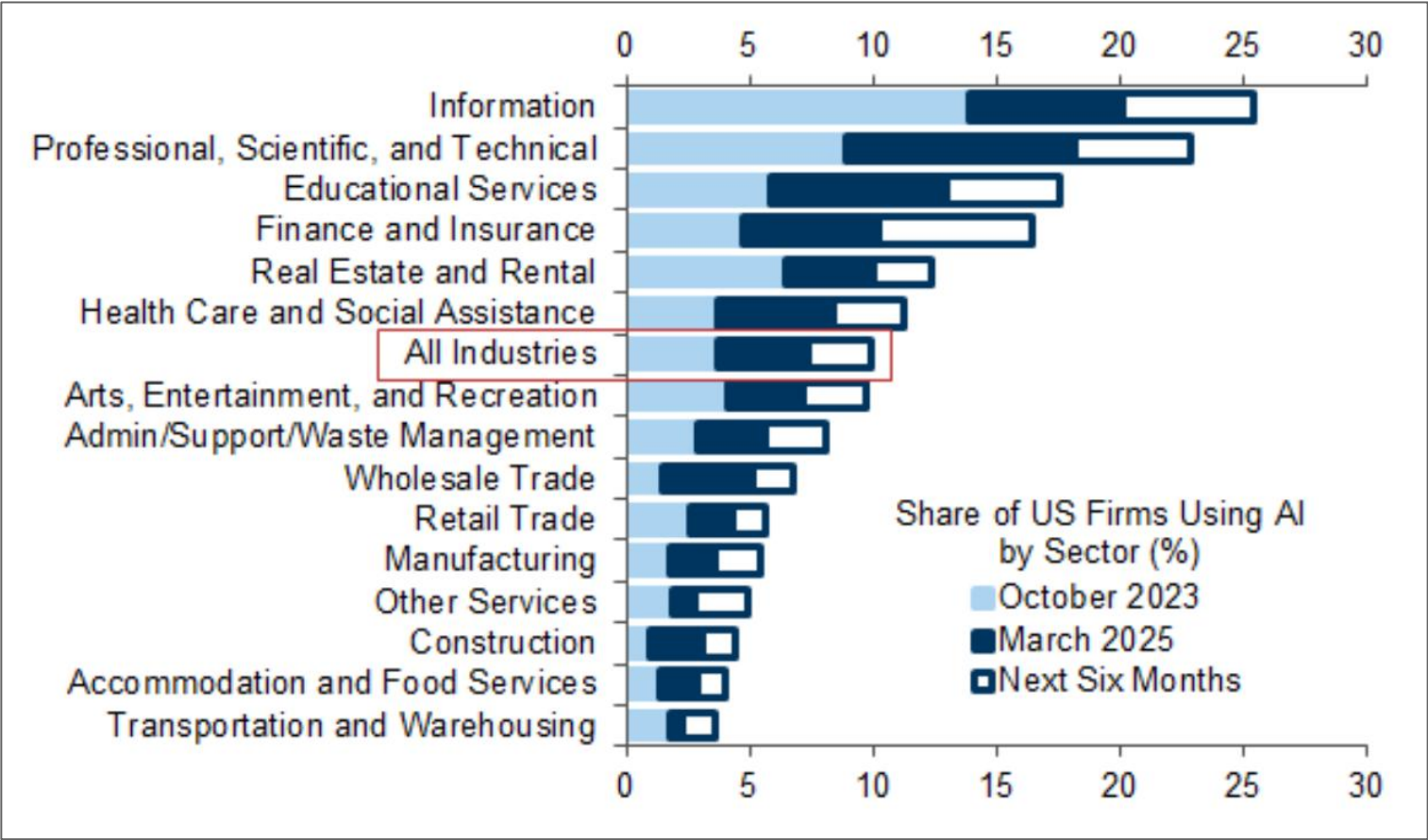
% of USA Firms Using AI – 3/25, per USA Census Bureau & Goldman Sachs Research



Note: Question asked was 'In the last six months, did this business use Artificial Intelligence (AI) in producing goods or services?' BTOS data are representative of all employer businesses in the USA economy, excluding farms. The BTOS sample consists of approximately 1.2MM businesses with biweekly data collection. Source: Census Bureau's BTOS (Business Trends & Outlook Survey) via Goldman Sachs Global Investment Research, '2025 Q1: Adoption Makes Modest Progress, Labor Impacts Still Negligible' (3/25)

# AI Adoption @ USA Firms = Rising...

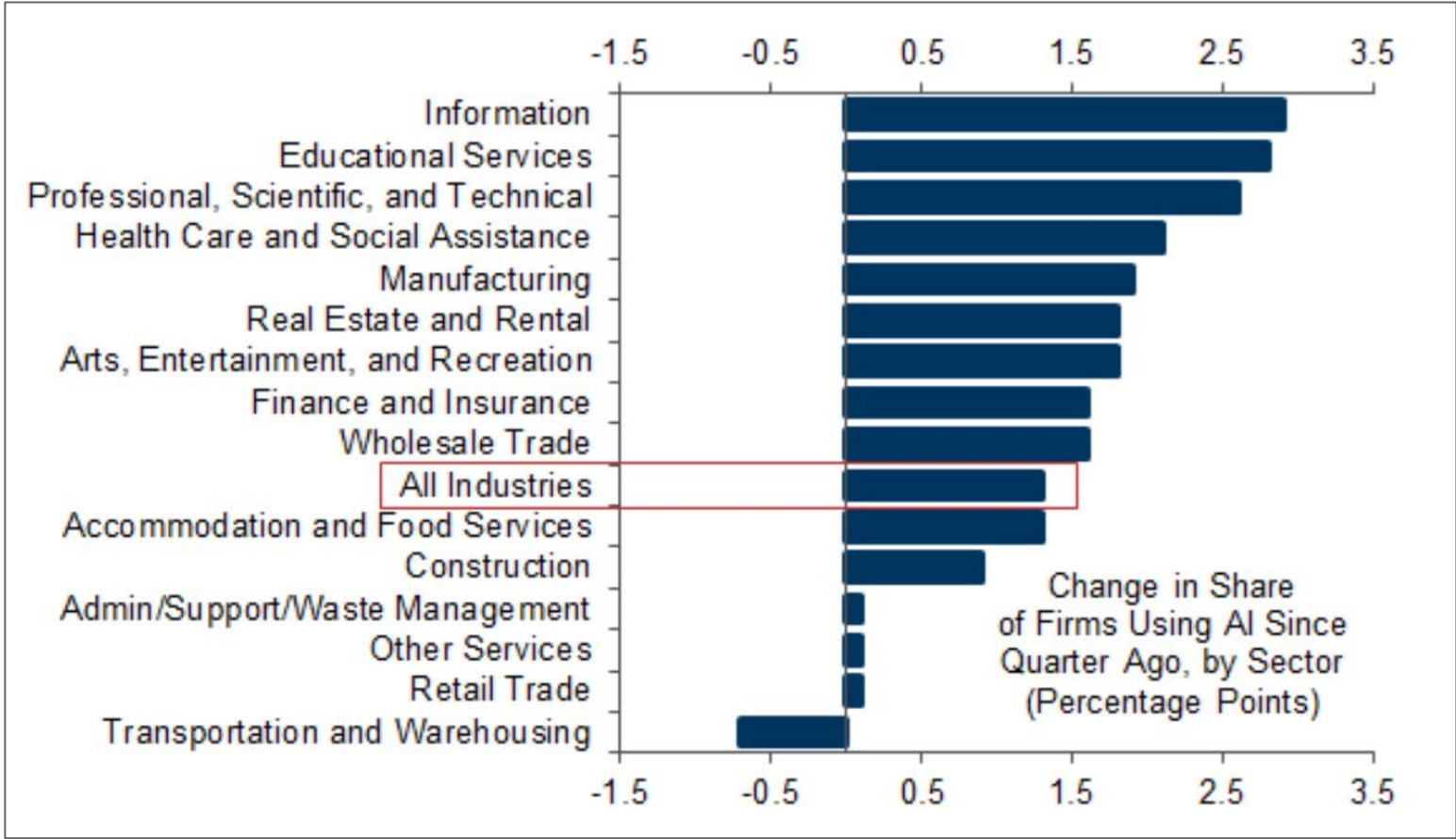
% of USA Firms Using AI – 3/25, per USA Census Bureau & Goldman Sachs Research



注意：提出的问题是“在过去六个月中，该企业是否在生产商品或服务中使用人工智能（AI）？”BTOS 数据代表美国经济中除农场外的所有雇主企业。BTOS 样本包括大约 120 万家企业，每两周收集一次数据。来源：美国人口普查局的 BTOS（商业趋势与展望调查），通过 Goldman Sachs Global Investment Research, “2025 年第一季度：采用取得适度进展，劳动力影响仍然可以忽略不计”（3/25）

...AI Adoption @ USA Firms =  
+21% Q/Q @ ~7% of Companies (Q1:25)

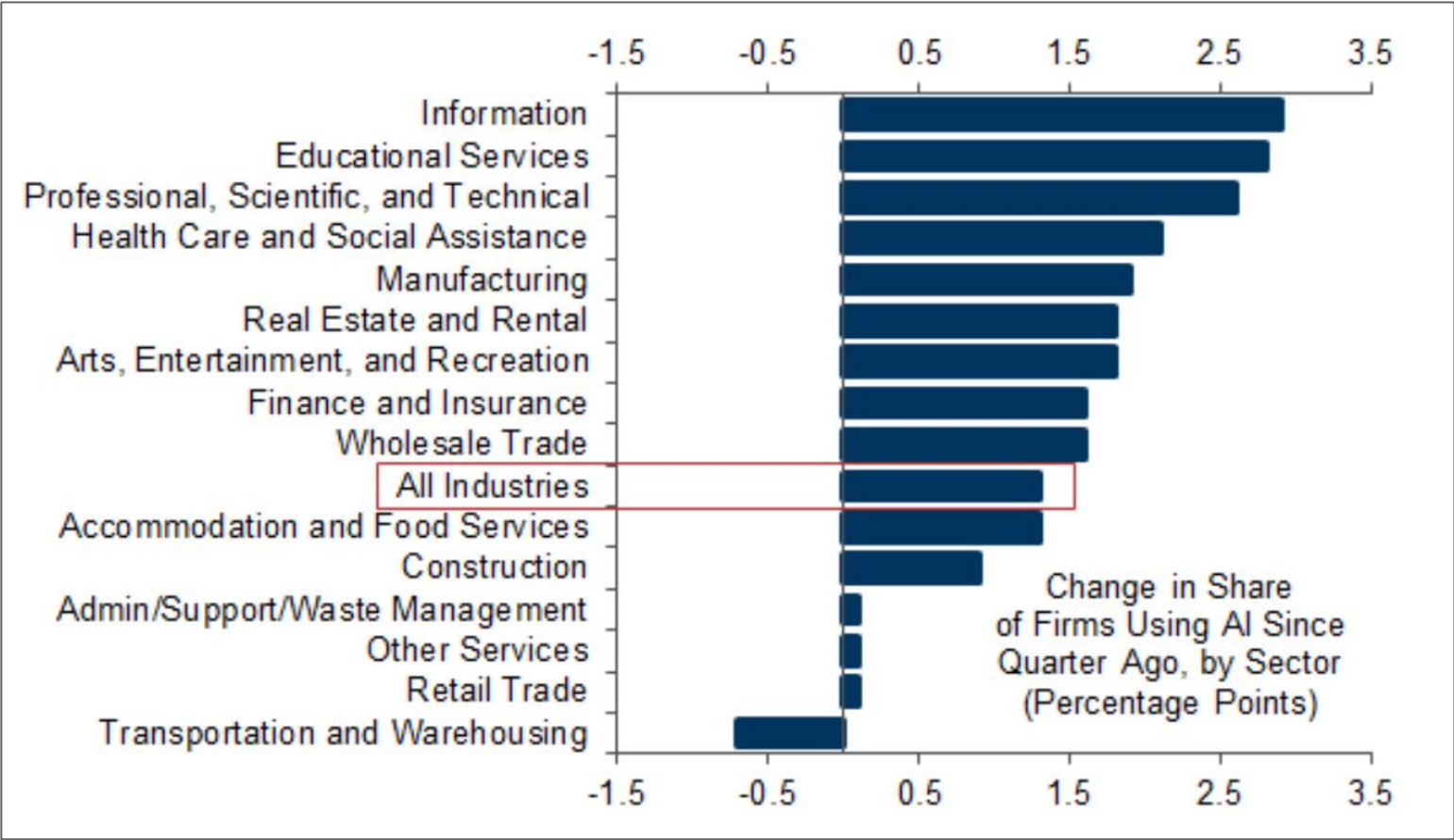
Change in % of USA Firms Using AI – Q4:24-Q1:25,  
per USA Census Bureau & Goldman Sachs Research



Note: Question asked was 'In the last six months, did this business use Artificial Intelligence (AI) in producing goods or services?' BTOS data are representative of all employer businesses in the USA economy, excluding farms. The BTOS sample consists of approximately 1.2MM businesses with biweekly data collection.  
Source: Census Bureau's BTOS (Business Trends & Outlook Survey) via Goldman Sachs Global Investment Research, '2025Q1: Adoption Makes Modest Progress, Labor Impacts Still Negligible' (3/25)

...AI Adoption @ USA Firms =  
+21% Q/Q @ ~7% of Companies (Q1:25)

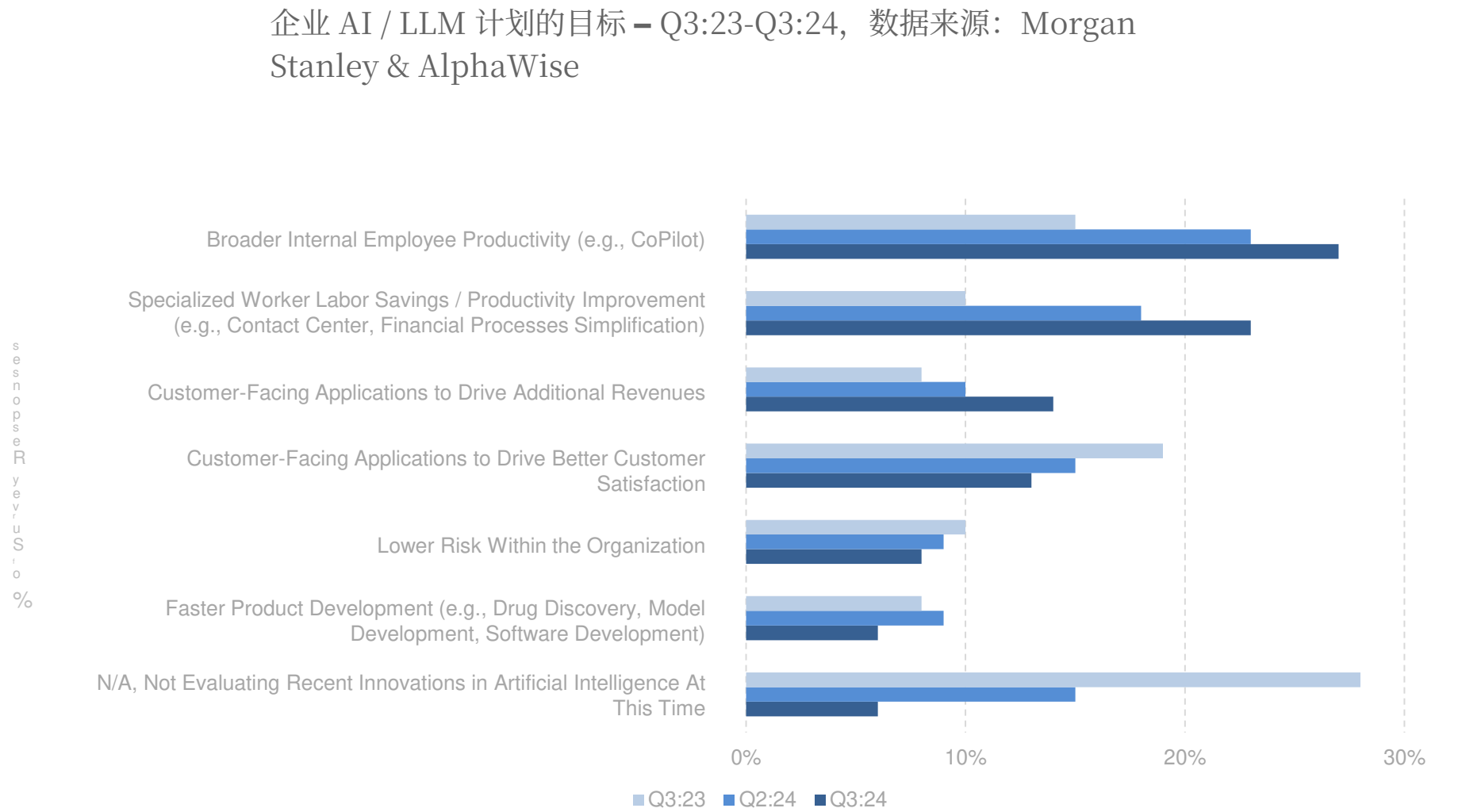
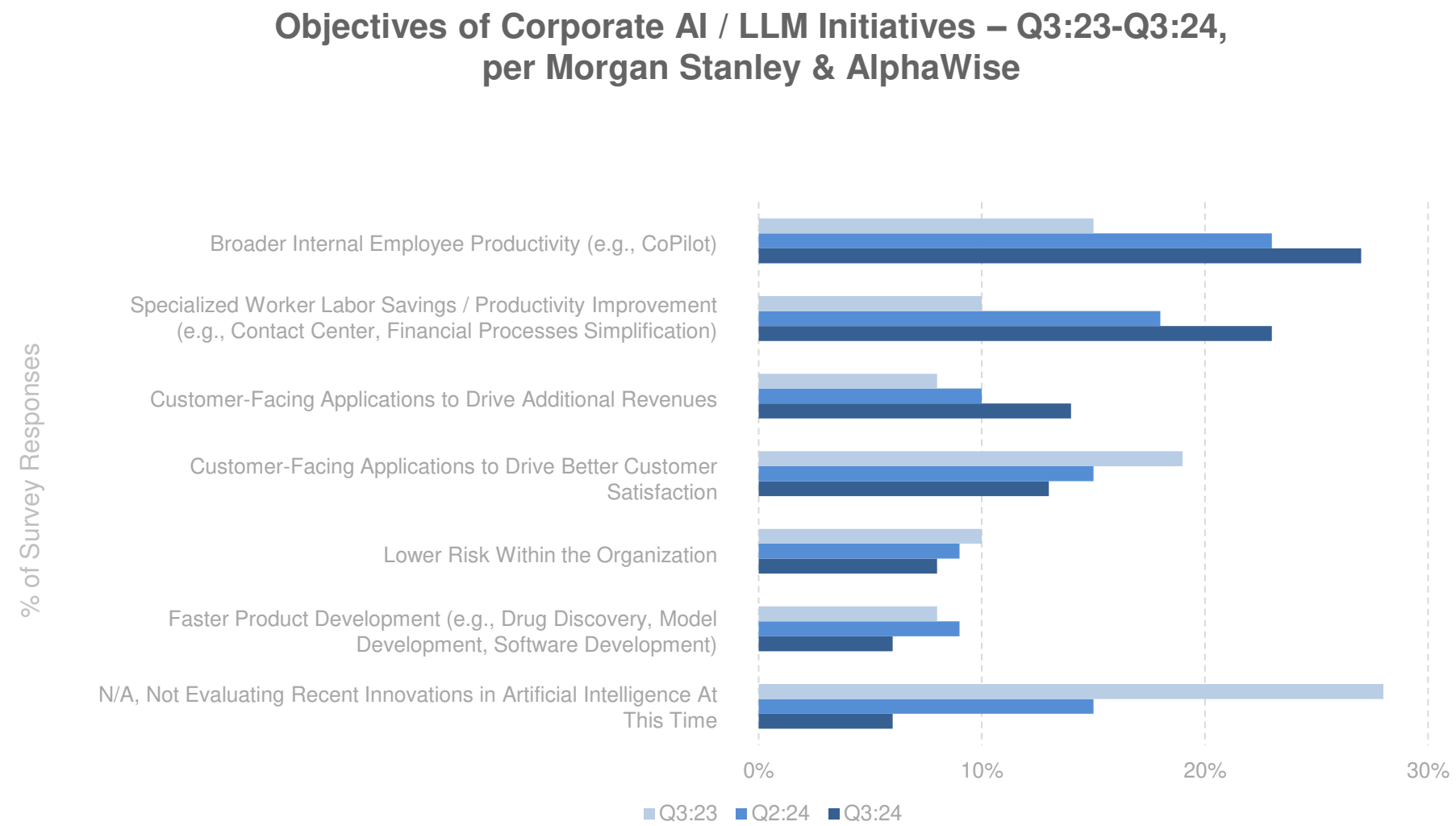
美国公司使用人工智能的百分比变化 – Q4:24-Q1:25，数据  
来源：美国人口普查局和高盛研究



注意：问题是“在过去的六个月里，这家企业在生产商品或提供服务时是否使用了人工智能（AI）？”BTOS 数据代表了美国经济中除农场外的所有雇主企业。BTOS 样本包括大约 120 万家企业，每两周收集一次数据。来源：美国人口普查局的 BTOS（商业趋势与展望调查），通过高盛全球投资研究，“2025Q1：采纳进展不大，劳动力影响仍然微不足道”（3/25）

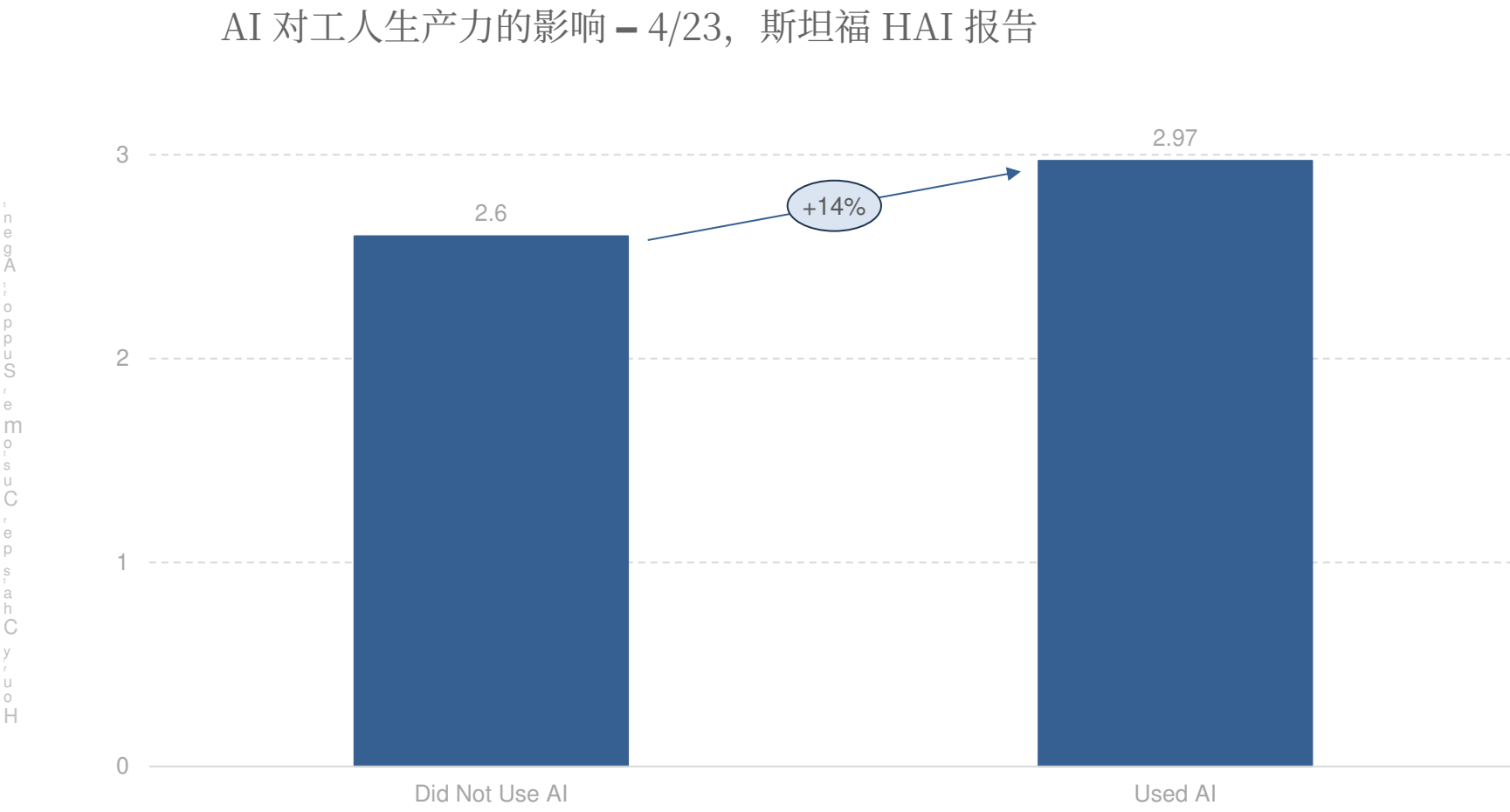
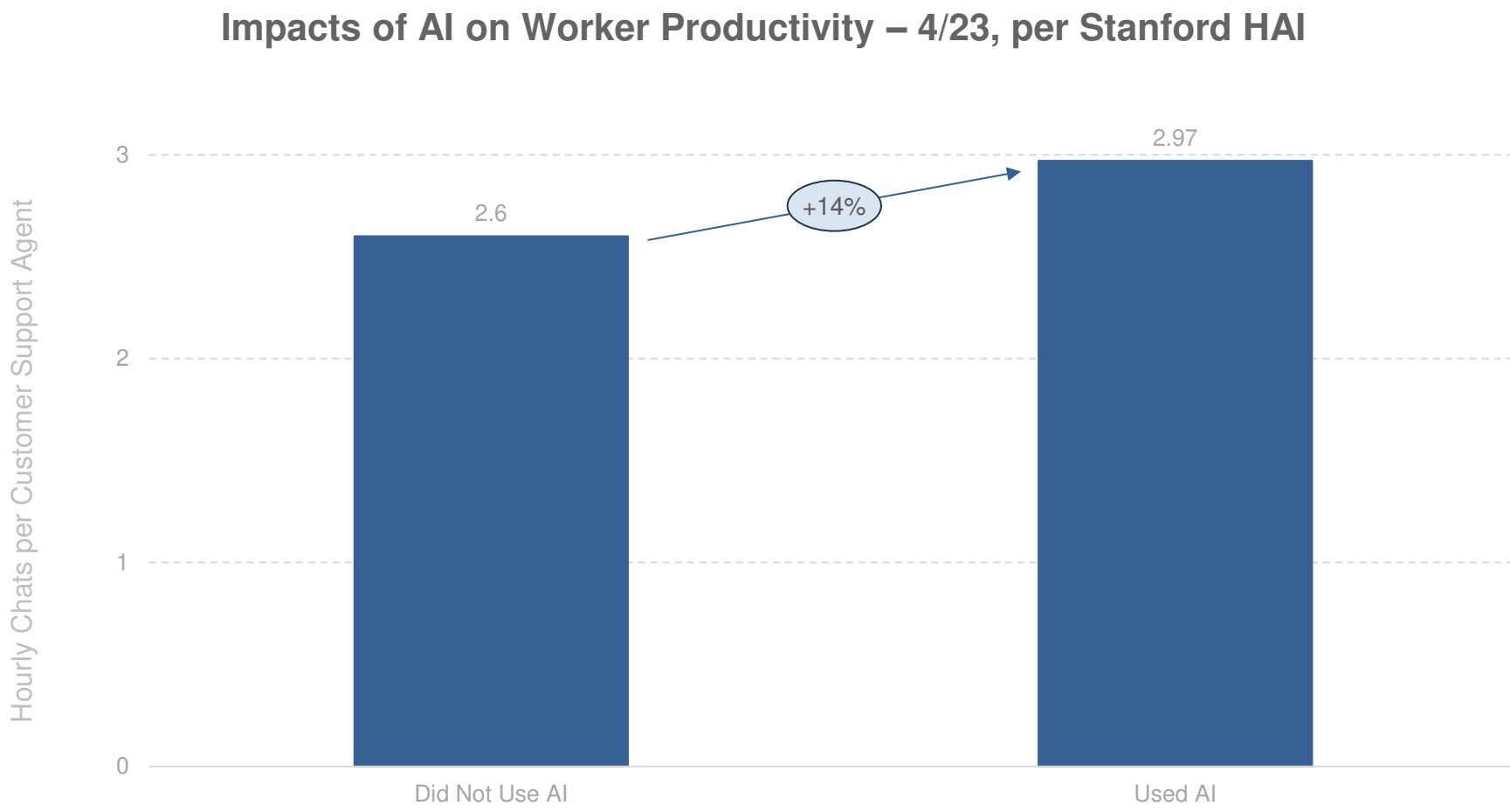
# AI Impact on Workforce = Employers Adopting AI to Drive Productivity Improvements

# AI 对劳动力的影响 = 雇主采用 AI 来提高生产力



AI Impact on Workforce = Seeing Productivity Gains, per Stanford HAI

AI 对劳动力的影响 = 斯坦福 HAI 报告显示生产力提升

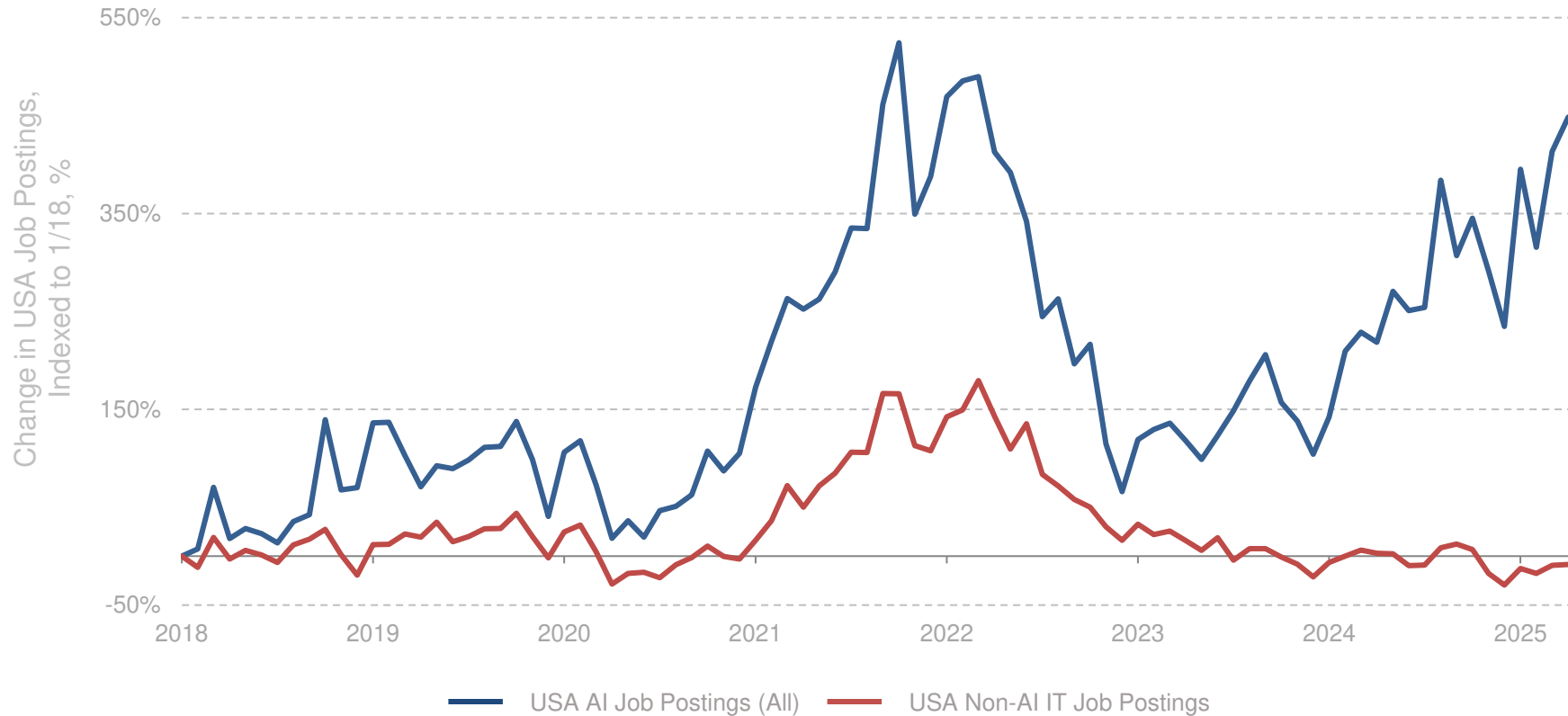


Note: Left chart: N = 5,179 customer support agents. Right chart: N = 1,018 scientists.  
Source: Erik Brynjolfsson et al., 'Generative AI at Work' (2/25) via Nestor Maslej et al., 'The AI Index 2025 Annual Report,' AI Index Steering Committee, Stanford HAI (4/25)

Note: Left chart: N = 5,179 customer support agents. Right chart: N = 1,018 scientists.  
Source: Erik Brynjolfsson et al., 'Generative AI at Work' (2/25) via Nestor Maslej et al., 'The AI Index 2025 Annual Report,' AI Index Steering Committee, Stanford HAI (4/25)

Employment Evolution – 1/18-4/25 =  
AI Job Postings +448% Over 7 Years While Non-AI IT Jobs -9%

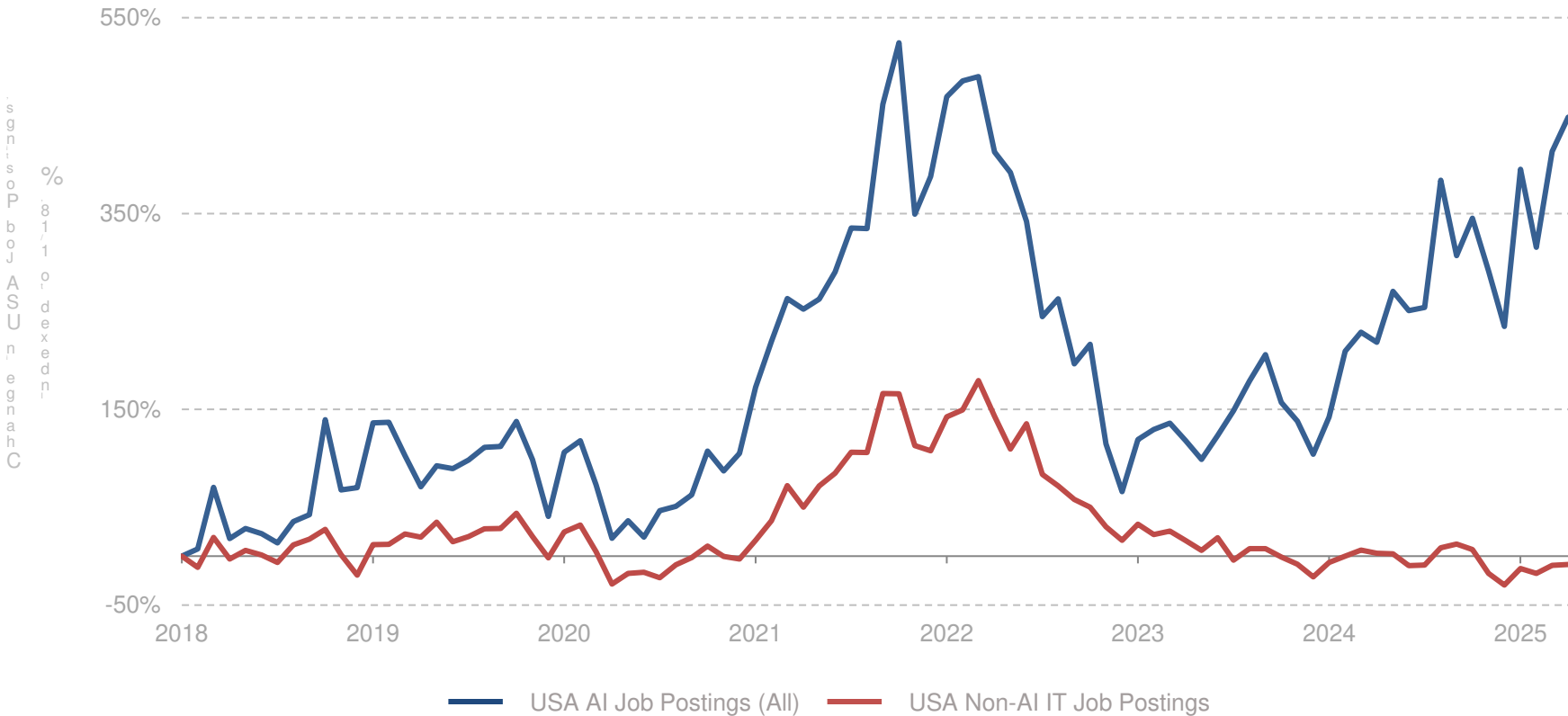
Change in USA AI & Non-AI IT Job Postings – 1/18-4/25,  
per University of Maryland & LinkUp



Note: 'AI Job' refers to a job posting that requires AI skills. AI skills requirement in job postings determined using University of Maryland's language processing model. USA-based jobs only. Figures are rounded. Source: University of Maryland's UMD-LinkUp AIMaps (in collaboration with Outrigger Group) (2/25)

Employment Evolution – 1/18-4/25 =  
AI Job Postings +448% Over 7 Years While Non-AI IT Jobs -9%

美国 AI 和非 AI IT 职位发布的变化 – 1/18-4/25，数据来源：马里，  
兰大学和 LinkUp

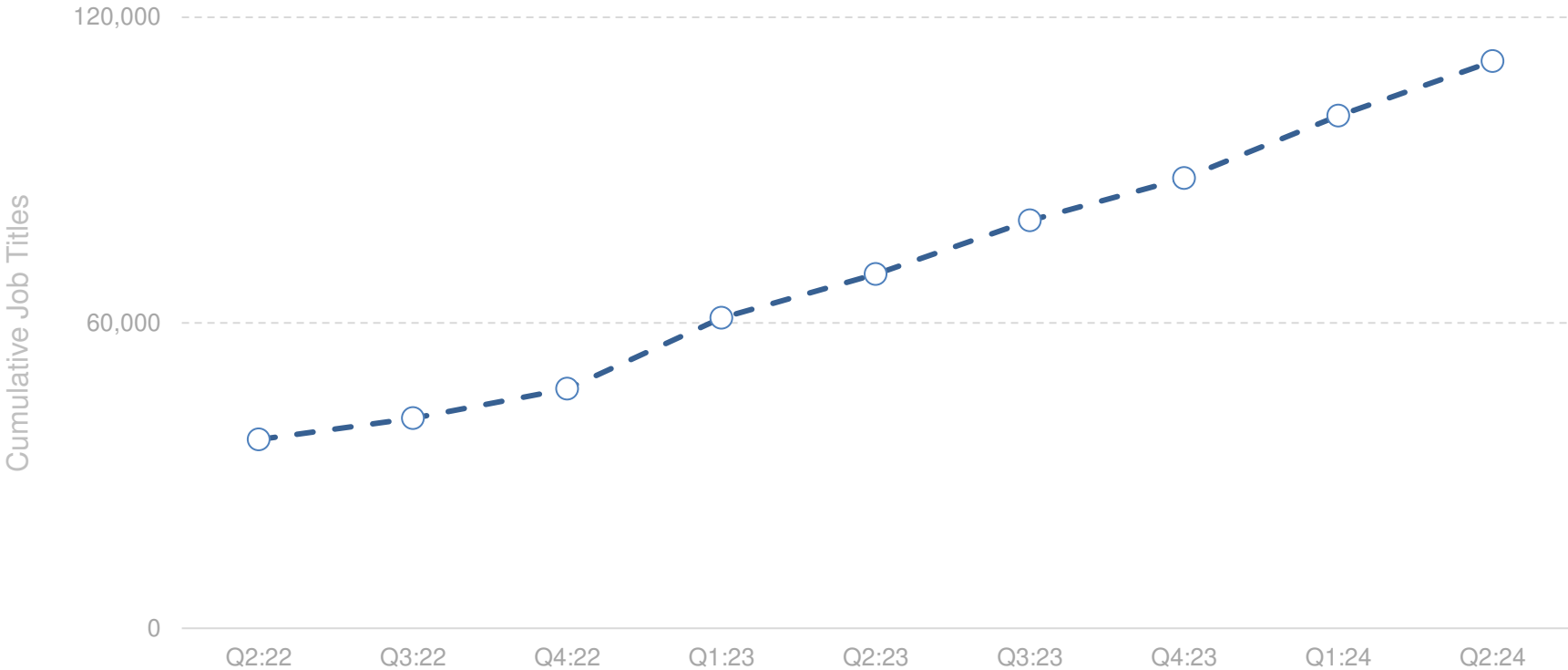


注意：“AI 职位”指的是需要 AI 技能的职位。职位发布中的 AI 技能要求由马里兰大学的语言处理模型确定。仅限位于 y of Maryland's language processing model. USA 美国的职位 only. Figures are rounded. Source: University of Maryland's UMD. 数字已四舍五入。来源：马里兰大学的 UMD-LinkUp AIMaps（与 Outrigger Group 合作）(2/25)



Employment Evolution – Q2:22-Q2:24 =  
AI-Related Job Titles +200% Over Two Years

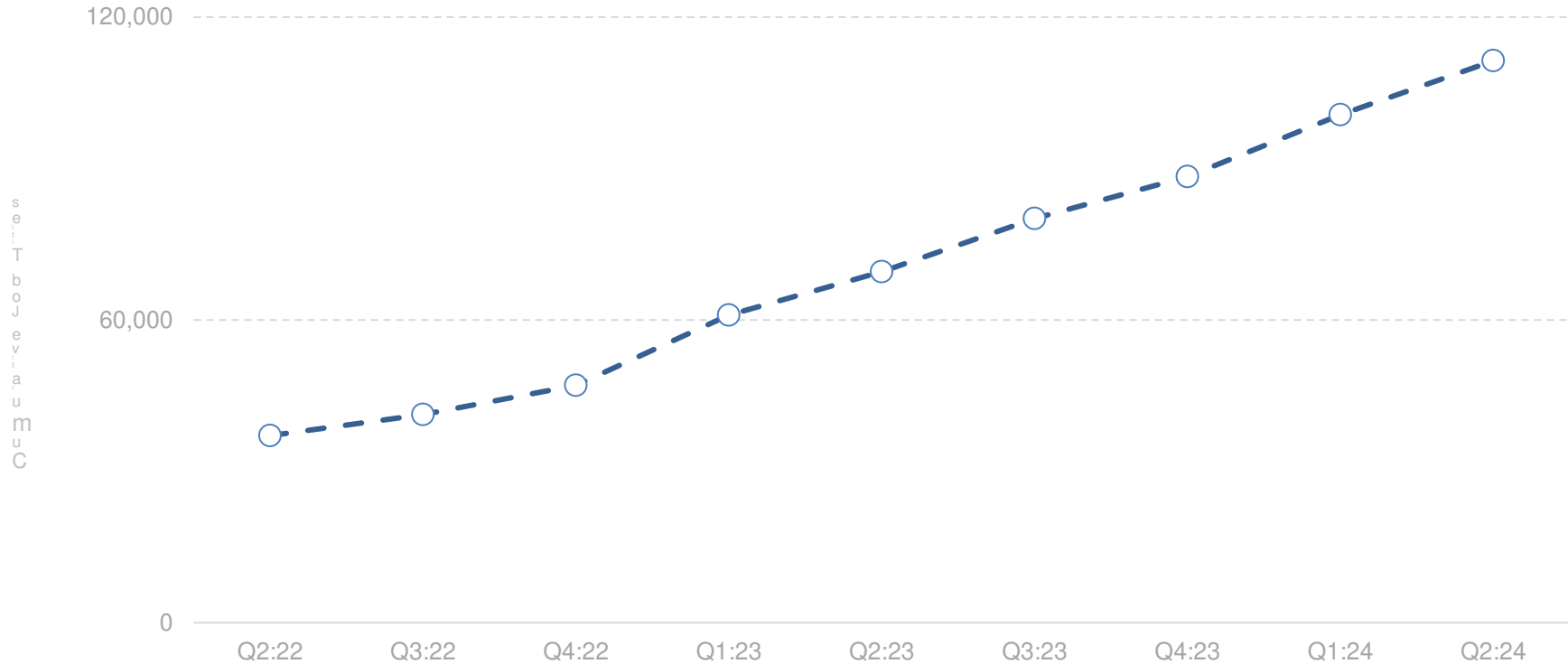
Cumulative # of New Global Job Titles With AI Terms Newly-Added – Q2:22-Q2:24,  
per ZoomInfo



*Note: The data in this report is sourced from ZoomInfo’s proprietary professional contacts database – a leading platform that detects more than 1.5MM personnel changes per day. To compile the trends in job titles, ZoomInfo’s data scientists analyzed announcements from hundreds of companies detailing their AI titles from 1/1/22 through 6/30/24. ZoomInfo’s database includes 100MM companies, 340MM professionals, & 11MM C-Suite leaders. Source: ZoomInfo (8/24)*

Employment Evolution – Q2:22-Q2:24 =  
AI-Related Job Titles +200% Over Two Years

Cumulative # of New Global Job Titles With AI Terms Newly-Added – Q2:22-Q2:24,  
per ZoomInfo

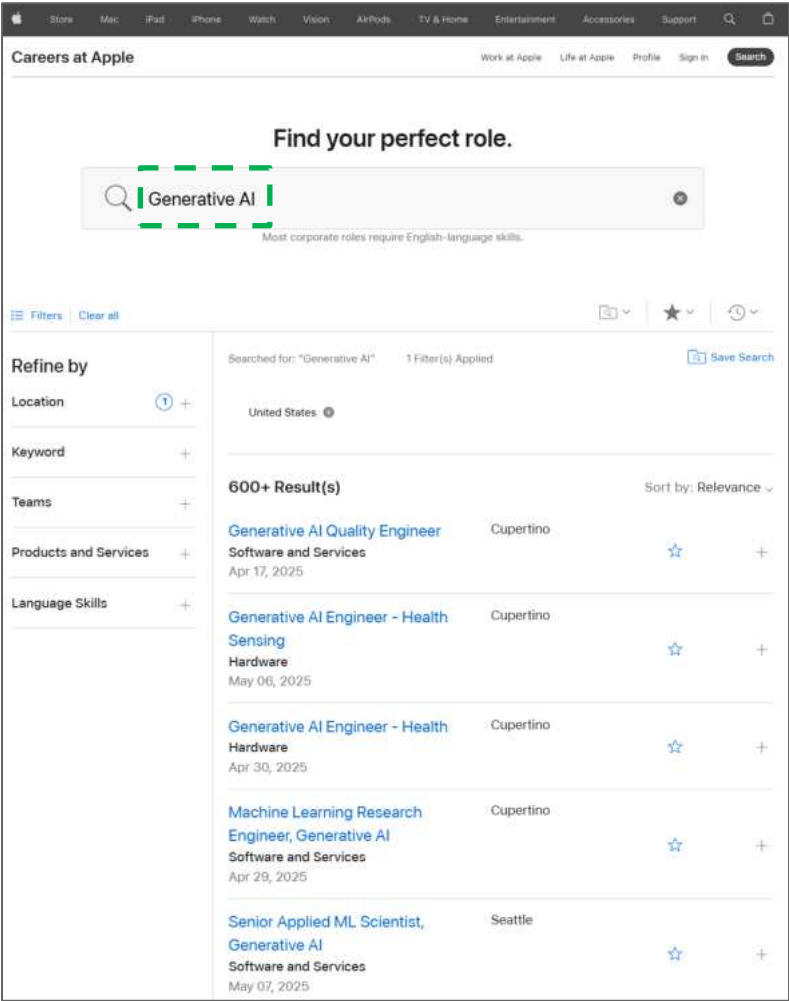


注意：本报告中的数据来源于 ZoomInfo 的专有专业联系人数据库 —— 一个领先的平台，每天检测超过 150 万的人事变动。为了整理职位头衔的趋势，ZoomInfo 的数据科学家分析了数百家公司的公告，详细说明了他们从 2022 年 1 月 1 日到 2024 年 6 月 30 日的人工智能职位。ZoomInfo 的数据库包括 1 亿家公司、3.4 亿专业人士和 1100 万 C 级领导。来源：ZoomInfo (8/24)



# Employment Evolution – Apple = 600+ Openings for Generative AI Jobs

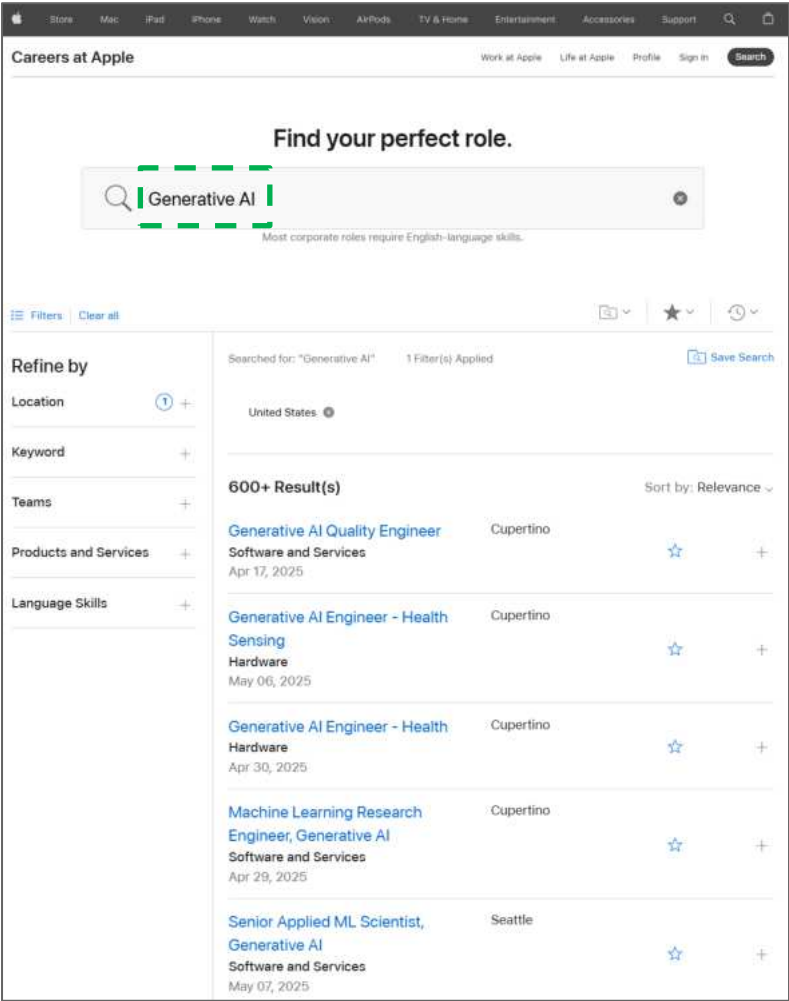
## Apple Job Postings Related to ‘Generative AI’ – 5/25



Source: Apple (4/25)

# 就业演变 – Apple =600+ 生成式 AI 职位空缺

## 与 “生成式 AI” 相关的 Apple 职位发布 – 5/25



Source: Apple (4/25)

### Example job description:

*As a member of the team you will be responsible for bringing innovative ideas and applying modern machine learning methods to solve problems that matter. From ideation to productization, you will participate in the full development cycle of core technologies, including handwriting and text recognition, handwriting synthesis, document understanding, freeform drawing recognition and generation. The ideal candidate should have experience in computer vision, speech recognition, deep learning, and/or other applications of machine learning systems.*

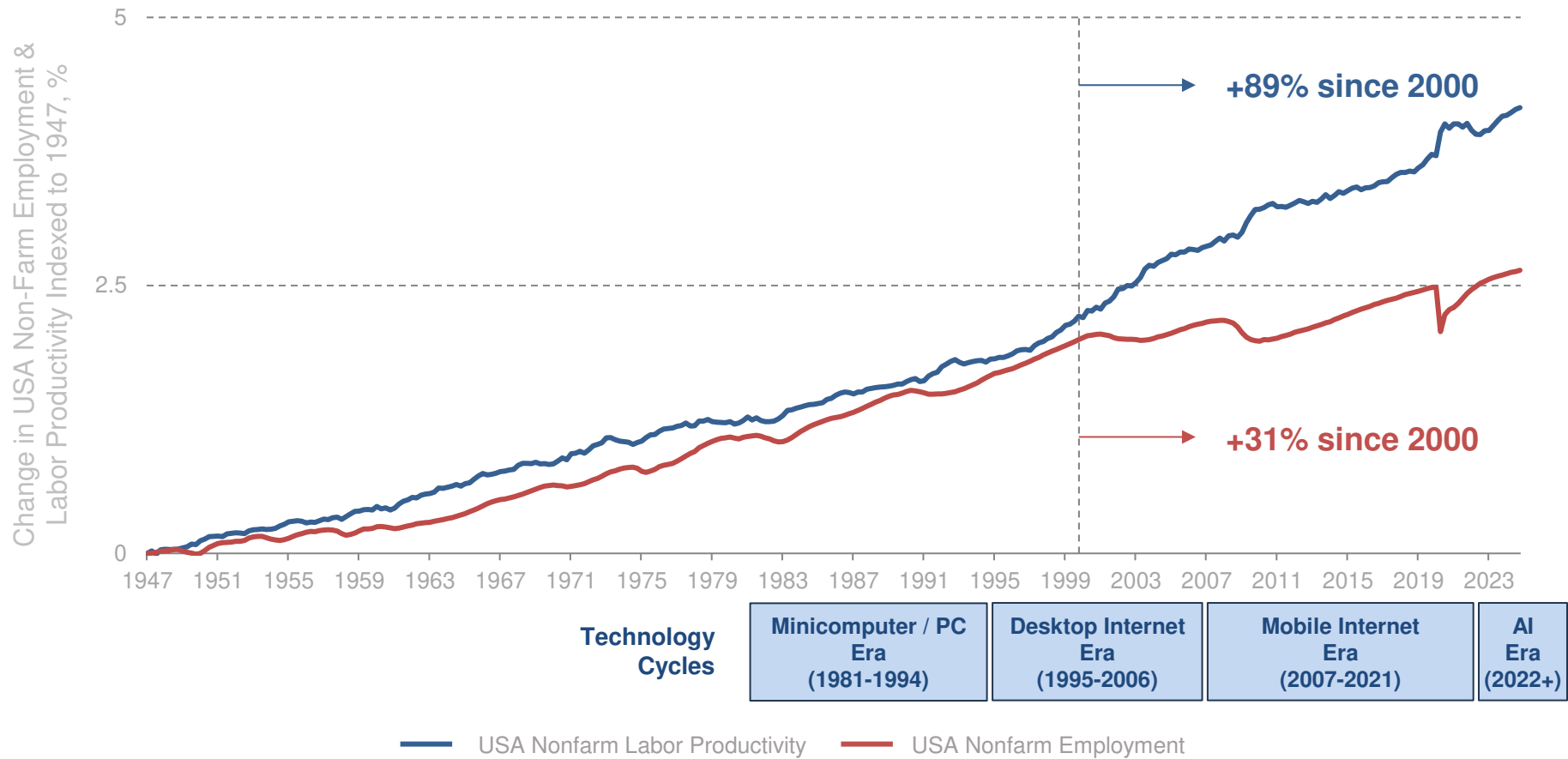
### 示例职位描述：

作为团队的一员，您将负责提出创新想法，并应用现代机器学习方法来解决重要问题。从构思到产品化，您将参与核心技术的完整开发周期，包括手写和文本识别、手写合成、文档理解、自由绘图识别和生成。理想的候选人应具有计算机视觉、语音识别、深度学习和 / 或机器学习系统其他应用方面的经验。

# USA Labor Productivity = Has Happened Alongside Job Growth Over Seventy-Seven Years

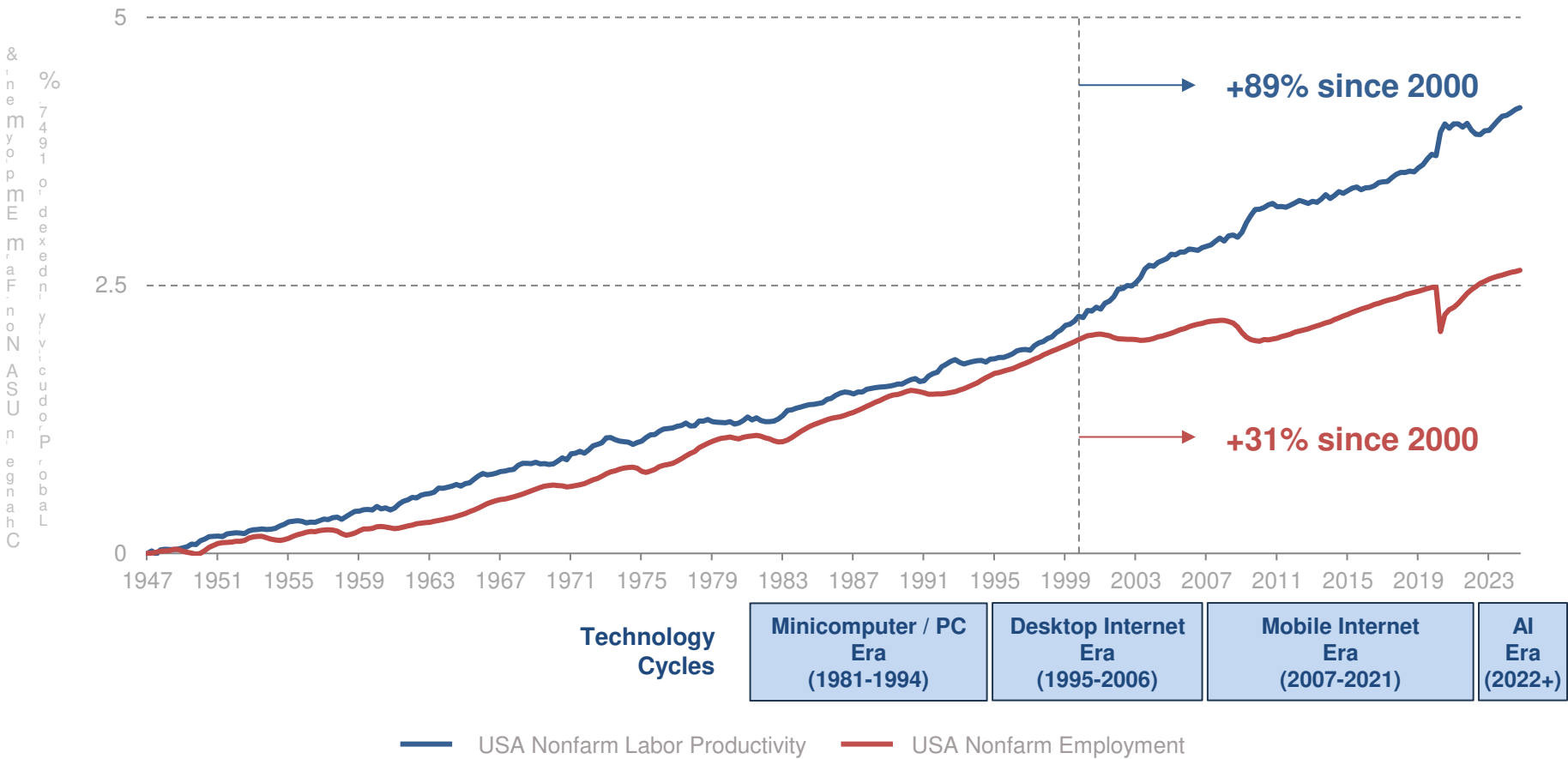
美国劳动生产率 = 在过去七十七年中与就业增长同时发生

Relative Change in USA Non-Farm Employment & Labor Productivity – 1947-2024,  
per Federal Reserve Bank of St. Louis



Note: Here we define the start of the PC Era as 1981 (launch of IBM PC). We define the start of the desktop internet era as 1995 (Netscape's IPO). We define the start of the mobile internet era as 2007 (the launch of Apple's iPhone). We define the start of the AI Era as 2022 (the public launch of ChatGPT). Source: Federal Reserve Bank of St. Louis (2024)

Relative Change in USA Non-Farm Employment & Labor Productivity – 1947-2024,  
per Federal Reserve Bank of St. Louis



注意：这里我们将 PC 时代的开始定义为 1981 年（IBM PC 的发布）。我们将桌面互联网时代的开始定义为 1995 年（Netscape 的首次公开募股）。我们将移动互联网时代的开始定义为 2007 年（Apple 的 iPhone 发布）。我们将人工智能时代的开始定义为 2022 年（ChatGPT 的公开发布）。资料来源：圣路易斯联邦储备银行（2024 年）

# AI In Workforce – NVIDIA = You're Not Going to Lose...Your Job to an AI...[But] to Somebody Who Uses AI

*All of you have heard a lot about [AI] job displacement. Every job will be affected. Some jobs will be lost, some jobs will be created, but every job will be affected. And immediately it is unquestionable, you're not going to lose a job – your job to an AI, but you're going to lose your job to somebody who uses AI...*

*...But let me give you the two extremes that you might want to consider as well. Computer technology, computer science has benefited about 30 million people. There are about 30 million people in the world who know how to program and use this technology to its extreme...*

*...The other eight, seven and a half billion people don't. I'll put on the table that, in fact, artificial intelligence is the greatest opportunity for us to close the technology divide. And let me prove it to you. You know, if we just look in this room, it's very unlikely that more than a handful of people know how to program with C++, and an equal number know how to program in C. And yet, 100 percent of you know how to program in AI. And the reason for that is because the AI will speak whatever language you wanted to speak...*

*...The number of people who are using ChatGPT and Gemini Pro and these AIs kind of demonstrate that, in fact, this is one of the easiest to use technologies in history...*

*...The other extreme that I will say is that, remember, we're – we have a shortage of labor. We have a shortage of workers.*

*We don't have an abundance of workers. We have a shortage of and for the very first time in history, we actually have – we can imagine the opportunity to close that gap to put 30-40 million workers back into the workforce that otherwise the world doesn't have. And so you could argue that artificial intelligence is probably our best way to increase the GDP, the global GDP, and so those are two other ways to look at it. In the meantime, I would recommend 100% of everybody you know take advantage of AI and don't be that person who ignores this technology.*

NVIDIA Co-Founder & CEO Jensen Huang @ Milken Institute Global Conference – 5/25

Source: Milken Institute (5/25)

# AI In Workforce – NVIDIA = 你不会因为人工智能而失去工作 ……[而是] 因为使用人工智能的人而失去工作

*All of you have heard a lot about [AI] job displacement. Every job will be affected. Some jobs will be lost, some jobs will be created, but every job will be affected. And immediately it is unquestionable, you're not going to lose a job – your job to an AI, but you're going to lose your job to somebody who uses AI...*

…… 但让我告诉你你可能想要考虑的两个极端。计算机技术，计算机科学已经使大约 3000 万人受益。世界上大约有 3000 万人知道如何编程并将其技术发挥到极致 ……

…… 其他 80 亿到 75 亿人不知道。我想说的是，事实上，人工智能是我们弥合技术鸿沟的最大机会。让我来证明给你看。你知道，如果我们只看这个房间，很可能只有少数人知道如何用 C++ 编程，也有相同数量的人知道如何用 C 编程。然而，你们 100% 的人都知道如何用人工智能编程。原因是因为人工智能会说你想说的任何语言 ……

… 使用 ChatGPT 和 Gemini Pro 以及这些 AI 的人数表明，事实上，这是历史上最容易使用的技术之一 ……

*...The other extreme that I will say is that, remember, we're – we have a shortage of labor. We have a shortage of workers.*

*We don't have an abundance of workers. We have a shortage of and for the very first time in history, we actually have – we can imagine the opportunity to close that gap to put 30-40 million workers back into the workforce that otherwise the world doesn't have. And so you could argue that artificial intelligence is probably our best way to increase the GDP, the global GDP, and so those are two other ways to look at it. In the meantime, I would recommend 100% of everybody you know take advantage of AI and don't be that person who ignores this technology.*

NVIDIA 联合创始人兼首席执行官黄仁勋 @ Milken Institute 全球会议 – 5/25

Source: Milken Institute (5/25)

Imagine, for a moment, how different your next week would look if there were no internet. Every facet of modern life – how we work, how we communicate, how we govern, and more – would likely be turned on its head. The internet has been woven into so many facets of life, big and small, that – for many – it is difficult to imagine a world without it.

In the next decade or two, imagining a world without AI will likely feel the same.

Artificial intelligence is reshaping the modern landscape at breakneck speed. What began as research has scaled into emerging core infrastructure across industries – powering everything from customer support to software development, scientific discovery, education, and manufacturing. This document has aimed to map the pace and breadth of AI’s expansion, with particular focus on usage trends, cost dynamics, infrastructure buildout, and early monetization models. The through-line is clear: AI is accelerating, touching more domains, and becoming more embedded in how work gets done.

Catalyzing this growth is the global availability of easy-to-use multimodal AI tools (like ChatGPT) on pervasive mobile devices, augmented by a steep decline in inference costs and an explosion in model availability. Both closed and open-source tools are now widely accessible and increasingly capable, enabling solo developers, startups, and enterprises alike to experiment and deploy with minimal friction. Meanwhile, large tech incumbents are weaving AI deeper into their products – rolling out copilots, assistants, and even agents that reframe how users engage with technology. Whether through embedded intelligence in SaaS or agentic workflows in consumer apps, the interface layer is being rewritten in real time.

On the compute side, investment continues to scale dramatically. Capital expenditures across major cloud providers, chipmakers, and hyperscalers have hit new highs, driven by the race to enable real-time, high-volume inference at scale. The investment is not just in chips, but also in new data centers, networking infrastructure, and energy systems to support growing demand. Whether this level of capital expenditure persists remains to be seen, but as AI moves closer to the edge – in vehicles, farms, labs, and homes – the distinction between digital and physical infrastructure continues to blur.

The global race to build and deploy frontier AI systems is increasingly defined by the strategic rivalry between the United States and China. While USA companies have led the charge in model innovation, custom silicon, and cloud-scale deployment to-date, China is advancing quickly in open-source development, national infrastructure, and state-backed coordination. Both nations view AI not only as an economic tailwind but also as a lever of geopolitical influence. These competing AI ecosystems are amplifying the urgency for sovereignty, security, and speed...

想象一下，如果没有互联网，你下周的生活会多么不同。现代生活的方方面面 – 我们的工作方式、沟通方式、管理方式等等 – 都可能会发生翻天覆地的变化。互联网已经融入了生活的方方面面，无论大小，以至于 – 对许多人来说 – ，很难想象一个没有互联网的世界。

在未来一二十年里，想象一个没有人工智能的世界可能会有同样的感觉。

人工智能正以惊人的速度重塑现代格局。最初的研究已经扩展成为跨行业的新兴核心基础设施 – 为从客户支持到软件开发、科学发现、教育和制造业等各个领域提供动力。本文旨在描绘人工智能扩张的速度和广度，特别关注使用趋势、成本动态、基础设施建设和早期货币化模式。贯穿始终的主线是：人工智能正在加速发展，触及更多领域，并越来越深入地融入到工作完成方式中。

促成这种增长的是在全球范围内普及的易于使用的多模态人工智能工具（如 ChatGPT ），这些工具可在普及的移动设备上使用，并辅以推理成本的大幅下降和模型可用性的爆炸式增长。封闭和开源工具现在都可以广泛使用，并且功能越来越强大，使独立开发人员、初创公司和企业都能以最小的摩擦进行实验和部署。与此同时，大型科技巨头正在将人工智能更深入地融入到他们的产品中 – 推出副驾驶、助手，甚至是重塑用户与技术互动方式的代理。无论是通过 SaaS 中的嵌入式智能，还是消费者应用程序中的代理工作流程，界面层都在实时重写。

在计算方面，投资持续大幅增加。主要云提供商、芯片制造商和超大规模企业的资本支出创下新高，这主要是由于大规模实现实时、大批量推理的竞赛所推动的。投资不仅限于芯片，还包括新的数据中心、网络基础设施和能源系统，以支持不断增长的需求。这种资本支出水平是否会持续下去还有待观察，但随着人工智能越来越接近边缘 – 在车辆、农场、实验室和家庭中 – 数字基础设施和物理基础设施之间的区别将继续变得模糊。

构建和部署前沿人工智能系统的全球竞赛越来越取决于美国和中国之间的战略竞争。虽然美国公司迄今为止在模型创新、定制芯片和云规模部署方面处于领先地位，但中国在开源开发、国家基础设施和国家支持的协调方面正在迅速发展。两国都将人工智能不仅视为经济推动力，而且视为地缘政治影响力的杠杆。这些相互竞争的人工智能生态系统正在增强对主权、安全和速度的迫切需求 ……



...In this environment, innovation is not just a business advantage; it is national posture.

As Microsoft Vice Chair and President Brad Smith recently noted,  
*Given the nature of technology markets and their potential network effects, this race between the U.S. and China for international influence likely will be won by the fastest first mover. Hence, the United States needs a smart international strategy to rapidly support American AI around the world...*  
*...The Chinese wisely recognize that if a country standardizes on China’s AI platform, it likely will continue to rely on that platform in the future. The best response for the United States is not to complain about the competition but to ensure we win the race ahead. This will require that we move quickly and effectively to promote American AI as a superior alternative. And it will need the involvement and support of American allies and friends.*

Lastly, AI is changing how we interact with the world around us. With affordable satellite connectivity expanding access to remote and underserved regions, the next wave of internet users will likely come online through AI-native experiences – skipping traditional app ecosystems and jumping straight into conversational, multimodal agents.

Similarly, AI uptake is accelerating in the workplace and has the potential to shape how people spend the one third of their lives at work. As usage patterns evolve and unit costs decline, we may be witnessing the early stages of an internet where intelligence is the default interface – accessible, contextual, and increasingly personal.

This is all amplified by the growing flow and transparency of information and capital – and the increasing examples of weaponization.

It comes at a time when global powers are more openly asserting autocracy-versus-democracy agendas. As technology and geopolitics increasingly intertwine, uncertainty is rising.

One thing is certain – it’s gametime for AI, and it’s only getting more intense... and the genie is not going back in the bottle.

...In 在这种环境下，创新不仅仅是一种商业优势，更是一种国家姿态。

正如微软副董事长兼总裁布拉德·史密斯最近指出的那样，鉴于技术市场的性质及其潜在的网络效应，美国和中国之间争夺国际影响力的竞赛很可能由最快的先行者赢得。因此，美国需要一项明智的国际战略，以迅速在全球范围内支持美国人工智能 …………… 中国人明智地认识到，如果一个国家以中国的 AI 平台为标准，那么它将来很可能会继续依赖该平台。美国最好的回应不是抱怨竞争，而是确保我们在未来的比赛中获胜。这将要求我们迅速有效地采取行动，以推广美国人工智能作为一种卓越的替代方案。这将需要美国盟友和朋友的参与和支持。

最后，人工智能正在改变我们与周围世界互动的方式。随着价格合理的卫星连接扩大了对偏远和服务欠缺地区的访问，下一波互联网用户可能会通过 AI 原生体验上线 – 跳过传统的应用程序生态系统，直接进入会话式多模态代理。

同样，人工智能在工作场所的应用正在加速，并有可能影响人们如何度过他们生命中三分之一的工作时间。随着使用模式的演变和单位成本的下降，我们可能正在见证互联网的早期阶段，在这个阶段，智能是默认的界面 – 可访问、情境化且日益个性化。

信息的流动和透明度以及资本的日益增长 – 以及越来越多的武器化例子，都放大了这一点。

目前，全球大国正在更加公开地宣扬专制与民主的议程。随着技术和地缘政治日益交织在一起，不确定性也在上升。

有一件事是肯定的 – it’s gametime for AI, and it’s only getting more intense... 而且精灵不会回到瓶子里。

BOND is a global technology investment firm that supports visionary founders throughout their entire life cycle of innovation and growth. BOND's founding partners have backed industry pioneers including Airbnb, AlphaSense, Applied Intuition, Canva, DocuSign, DoorDash, KoBold Metals, Meta (Facebook), Instacart, Peloton, Plaid, Revolut, Slack, Spotify, Square, Stripe, Twitter, Uber, & VAST Data.

This document, including the information contained herein, has been compiled for informational purposes only & does not constitute an offer to sell or a solicitation of an offer to purchase any security. Such offer or solicitation shall only be made pursuant to offering documents related to such security, if any.

The document relies on data + insights from a wide range of sources, including public + private companies, market research firms + government agencies. We cite specific sources where data is public; the document is also informed by non-public information + insights. We disclaim any + all warranties, express or implied, with respect to the document.


We do not take a view on the relative veracity of data sources or inconsistencies expressed or implied therefrom. No document content should be construed as professional advice of any kind (including legal or investment advice).


We may post updates, revisions, or clarifications of this document on BOND’s website (www.bondcap.com).

BOND owns or has owned significant equity positions in certain of the companies referenced in this document.

Data Provided Under License to BOND

yipitDATA

 SensorTower

 SimilarWeb

BOND 是一家全球技术投资公司，在创新和增长的整个生命周期中为有远见的创始人提供支持。BOND 的创始合伙人支持了包括 Airbnb、AlphaSense、Applied Intuition、Canva、DocuSign、DoorDash、KoBold Metals、Meta (Facebook)、Instacart、Peloton、Plaid、Revolut、Slack、Spotify、Square、Stripe、Twitter、Uber 和 VAST Data 等行业先驱。

本文档（包括其中包含的信息）仅为提供信息而编写，不构成出售要约或购买任何证券的要约邀请。此类要约或邀请应仅根据与此类证券相关的要约文件（如有）发出。

本文档依赖于来自广泛来源的数据 + 见解，包括公共 + 私营公司、市场调查公司 + 政府机构。我们在数据公开的地方引用具体来源；本文档也参考了非公开信息 + 见解。我们不承担与本文档相关的任何 + 所有明示或暗示的保证。


我们不对数据来源的相对真实性或由此表达或暗示的不一致之处持任何观点。任何文档内容都不应被解释为任何形式的专业建议（包括法律或投资建议）。


我们可能会在 BOND 的网站 (www.bondcap.com) 上发布本文档的更新、修订或澄清。

BOND 拥有或曾经拥有本文档中引用的某些公司的重要股权。

Data Provided Under License to BOND

yipitDATA

 SensorTower

 SimilarWeb