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

趋势 – 人工智能 (AI)2025 年 5 月 30 日

玛丽 · 梅考克 / 杰伊 · 西蒙斯 / 蔡大元 / 亚历山大 · 克雷

BOND

Context

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.

Context

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.

与人工智能技术演变相关的变化速度和范围确实是无与伦比的，这一点有数据支持。本文件充满了上升的用户、使用和收入图表，这些图表往往由也呈上升趋势的支出图表所支持。

创作者 / 投注者 / 消费者正在利用全球互联网轨道，这些轨道通过连接设备可被 55 亿公民使用；不断增长的数字数据集，这些数据集已经持续了三十多年；突破性的大型语言模型（ LLMs ），这些模型在 2022 年 11 月发布 OpenAI 的 ChatGPT 后实际上获得了自由，其极其易于使用 / 快速的用户界面。

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.

我们文件的提纲在下一页，随后是十一个图表，这些图表有助于说明接下来的观察结果。

我们希望这份汇编能够增加关于正在发生的变革广度的讨论 – 技术 / 金融 / 社会 / 物理 / 地缘政治。毫无疑问，随着我们所有人努力适应这场不断发展的旅程，人们（和机器）将在这几点上取得进步，因为知识 – 及其分布 – 正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.

BOND 2

Outline

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

Outline

		#
1	似乎变化正在比以往任何时候都快？是的，就是这样	9-51
2	AI 用户 + 使用 + 资本支出增长 = 前所未有的	52-128
3	AI 模型计算成本高 / 上升 + 每令牌推理成本下降 = 性能趋同 + 开发者使用量上升	129-152
4	AI 使用 + 成本 + 损失增长 = 前所未有的	153-247
5	人工智能变现威胁 = 竞争加剧 + 开源势头 + 中国崛起	248-298
6	AI 与物理世界加速 = 快速 + 数据驱动	299-307
7	全球互联网用户由 AI 驱动迅速增长 = 前所未有的增长	308-322
8	AI & 工作进化 = 真实的 + 快速的	323-336

Charts Paint Thousands of Words…

1

Seem Like Change Happening Faster Than Ever? Yes, It Is

Developers in Leading Chipmaker's Ecosystem

Source: Leading Chipmaker

Details on Page 38

2

AI User + Usage + CapEx Growth = Unprecedented

Leading USA-Based LLM Users

Source: Company disclosures

Details on Page 55

2.1

AI User + Usage + CapEx Growth = Unprecedented

Internet vs. Leading USA-Based LLM: Total Current Users Outside North America

Note: LLM data is for monthly active mobile app users. App not available in select countries, including China and Russia, as of 5/25.
Source: United Nations / International Telecommunications Union (3/25), Sensor Tower (5/25)

Details on Page 56

2.2

AI User + Usage + CapEx Growth = Unprecedented

Big Six USA Technology Company CapEx*

*Apple, NVIDIA, Microsoft, Alphabet, Amazon (AWS only), & Meta Platforms
Source: Capital IQ (3/25), Morgan Stanley

Details on Page 97

图表描绘千言万语 …

1

似乎变化比以往任何时候都快？是的，就是这样

领先芯片制造商生态系统中的开发者

Source: Leading Chipmaker

Details on Page 38

2

AI 用户 + 使用 + 资本支出增长 = 前所未有

领先的美国 LLM 用户

Source: Company disclosures

Details on Page 55

2.1

AI 用户 + 使用 + 资本支出增长 = 前所未有

Internet vs. Leading USA-Based LLM: Total Current Users Outside North America

注意：LLM 数据是指月活跃移动应用用户。截至 5 月 25 日，该应用在某些国家不可用，包括中国和俄罗斯。来源：联合国 / 国际电信联盟（3 月 25 日），Sensor Tower （5 月 25 日）年数

Details on Page 56

2.2

AI 用户 + 使用 + 资本支出增长 = 前所未有

Big Six USA Technology Company CapEx*

*Apple, NVIDIA, Microsoft, Alphabet, Amazon (AWS only), & Meta Platforms
Source: Capital IQ (3/25), Morgan Stanley

Details on Page 97

BOND 4

…Charts Paint Thousands of Words…

3

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

Cost of Key Technologies Relative to Launch Year

Note: Per-token inference costs shown.
Source: Richard Hirsh; John McCallum; OpenAI

Details on Page 138

4

AI Usage + Cost + Loss Growth = Unprecedented

Leading USA-Based AI LLM Revenue vs. Compute Expense

Note: Figures are estimates.
Source: The Information, public estimates

Details on Page 173

5

AI Monetization Threats = Rising Competition + Open-Source Momentum + China's Rise

Leading USA LLMs vs. China LLM Desktop User Share

Note: Data is non-deduped. Share is relative, measured across six leading global LLMs.
Source: YipitData (5/25)

Details on Page 293

5.1

AI Monetization Threats = Rising Competition + Open-Source Momentum + China's Rise

China vs. USA vs. Rest of World Industrial Robots Installed

Note: Data as of 2023.
Source: International Federation of Robotics

Details on Page 289

… 图表描绘千字 …

3

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

Cost of Key Technologies Relative to Launch Year

Note: Per-token inference costs shown.
Source: Richard Hirsh; John McCallum; OpenAI

Details on Page 138

4

AI 使用 + 成本 + 损失增长 = 前所未有

Leading USA-Based AI LLM Revenue vs. Compute Expense

Note: Figures are estimates.
Source: The Information, public estimates

Details on Page 173

5

AI Monetization Threats = Rising Competition + Open-Source Momentum + China's Rise

Leading USA LLMs vs. China LLM Desktop User Share

Note: Data is non-deduped. Share is relative, measured across six leading global LLMs.
Source: YipitData (5/25)

Details on Page 293

5.1

AI Monetization Threats = Rising Competition + Open-Source Momentum + China's Rise

中国 vs. 美国 vs. 其他地区工业机器人安装

Note: Data as of 2023.
Source: International Federation of Robotics

Details on Page 289

…Charts Paint Thousands of Words

6

AI & Physical World Ramps = Fast + Data-Driven

A Ride Share vs. Autonomous Taxi Provider, San Francisco Operating Zone Market Share

Source: YipitData (4/25)

Details on Page 302

7

Global Internet User Ramps Powered by AI from Get-Go = Growth We Have Not Seen Likes of Before

Leading USA-Based LLM App Users by Region

Note: Region definitions per World Bank definitions. China not included in East Asia figures. Data for standalone app only. Source: Sensor Tower (5/25)

Details on Page 315

8

AI & Work Evolution = Real + Rapid

USA IT Jobs – AI vs. Non-AI

Source: University of Maryland's UMD-LinkUp AIMaps (in collaboration with Outrigger Group) (5/25)

Details on Page 332

… 图表描绘千字万言

6

AI & Physical World Ramps = Fast + Data-Driven

A Ride Share vs. Autonomous Taxi Provider, San Francisco Operating Zone Market Share

Source: YipitData (4/25)

Details on Page 302

7

Global Internet User Ramps Powered by AI from Get-Go = Growth We Have Not Seen Likes of Before

Leading USA-Based LLM App Users by Region

Note: Region definitions per World Bank definitions. China not included in East Asia figures. Data for standalone app only. Source: Sensor Tower (5/25)

Details on Page 315

8

AI & Work Evolution = Real + Rapid

USA IT Jobs – AI vs. Non-AI

Source: University of Maryland's UMD-LinkUp AIMaps (in collaboration with Outrigger Group) (5/25)

Details on Page 332

BOND 6

Overview…

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…

Overview…

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

谷歌的创始使命（1998 年）是 “ 组织全球信息并使其普遍可访问、有用 ”。阿里巴巴的创始使命（1999 年）是 “ 让天下没有难做的生意 ”。脸书的创始使命（2004 年）是 “ 赋予人们分享的力量，让世界更加开放和连接 ”。

快进到今天，全球组织化、连接化和可访问的信息正在被人工智能、加速计算能力和半无国界的资本超级驱动 …… 所有这些都正在推动巨大的变革。

运动为人工智能的持续改进提供了一个很好的类比。随着运动员继续让我们惊叹并打破纪录，他们的才能正越来越多地通过更好的数据 / 输入 / 训练得到增强。对于企业来说，情况也是如此，计算机正在摄入海量数据以变得更聪明和更具竞争力。大型模型的突破、每 token 成本下降、开源的普及和芯片性能的改进正在使新技术进步越来越强大、可访问和具有经济可行性。

OpenAI 的 ChatGPT – 基于用户 / 使用情况 / 营收指标 – 是历史上最大的 ‘ 一夜成名 ’ 成功 (成立九年后)。人工智能的使用在消费者、开发者、企业和政府中激增。与互联网 1.0 革命 – 技术始于美国并逐步全球扩散 –ChatGPT 一次性登上了世界舞台，同时在大多数全球地区同步增长。

与此同时，平台既得利益者和新兴挑战者正竞相建设和部署人工智能基础设施的下一层：代理界面、企业副驾驶、现实世界自主系统和国有模型。

人工智能、计算基础设施和全球互联的快速进步正在从根本上重塑工作方式、资本部署方式以及公司和国家层面的领导力定义 – 。

与此同时，全球大国正在经历领导力演变，每个国家都在挑战其他国家的竞争和比较优势。我们看到世界上最有权势的国家因不同程度的经济 / 社会 / 领土野心而备受激励 ……

…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.

…Overview

… 越来越明显，两大强大力量 – 技术和地缘政治 – 正在交织在一起。Andrew Bosworth （Meta Platforms 首席技术官），在一次 ‘可能’ 播客中描述了人工智能的当前状态为 我们的太空竞赛 和 我们正在讨论的人，尤其是中国，都非常有能力 …… 秘密很少。而且就是进步。而且你想确保你永远不落后。

现实是，人工智能的领导地位可能会带来地缘政治的领导地位 – 而不是反过来。

这种状况带来了巨大的不确定性 …… 但它又把我们带回了我们最喜欢的一句引言 —— 从统计学上来说，世界并不经常结束，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会出错，但令人兴奋的是考虑什么可能会成功。一次又一次，乐观的理由是人们可以做出的最佳赌注之一。看着人工智能为你工作所带来的魔力，感觉就像电子邮件和网络搜索 – 这些从根本上改变了我们世界的科技。人工智能带来的更好 / 更快 / 更便宜的影响似乎同样神奇，但甚至更快。

毫无疑问，这些也是危险和不确定的时期。但人工智能的长期乐观论基于这样一种观点，即激烈的竞争和创新 … 日益 - 可获取的计算 … 迅速 - 全球人工智能 - 融合技术 … 以及深思熟虑和计算的领导力可以培养足够的恐惧和尊重，进而可能导致相互确保摧毁。

对于一些人来说，人工智能的演变将导致向下的竞争；对于另一些人来说，它将导致向上的竞争。资本主义和创造性破坏的投机和狂热力量是地壳般的。不可否认的是，‘游戏开始了’，尤其是在美国和中国以及科技巨头奋勇前进的情况下。

在本文件中，我们分享了第三方使用他们认为有效的方法论的数据 / 研究 / 基准，我们感谢许多人在这独特而动态的时期努力说明趋势。我们的目标是补充讨论。

Outline

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

Outline

1. 似乎变化正在比以往任何时候都快？是的，就是这样
2. AI 用户 + 使用 + 资本支出增长 = 前所未有的
3. AI 模型计算成本高 / 上升 + 每令牌推理成本下降 = 性能趋同 + 开发者使用量上升
4. AI 使用 + 成本 + 损失增长 = 前所未有的
5. 人工智能变现威胁 = 竞争加剧 + 开源势头 + 中国崛起
6. AI 与物理世界加速 = 快速 + 数据驱动
7. 全球互联网用户由 AI 驱动迅速增长 = 我们从未见过的增长
8. AI & 工作进化 = 真实的 + 快速的

Technology Compounding =

Numbers Behind The Momentum

技术复合 =

Numbers Behind The Momentum

BOND 10

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

千余年技术复合效应 → 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，每麦迪逊计划

注意：图表以 2011 年 'GK$' 计算的万亿美元实际 GDP 表示，采用对数刻度。GK$ （知识总美元）是一个非正式术语，用于估算特定见解、想法或专有知识的潜在商业价值。它反映了如果有效应用，即使尚未产生收入，该知识可能的价值。来源：微软，《治理人工智能：未来蓝图》，微软报告（5/23）；数据来自麦迪逊计划 & 我们的世界 在数据

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

Computing Cycles Over Time – 1960s-2020s, per Morgan Stanley

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)

… 技术复合增长超过五十多年 = 更好 + 更快 + 更便宜 → 更多

Computing Cycles Over Time – 1960s-2020s, per Morgan Stanley

注意：2000 年的 PC 单位。2005 年的台式机互联网用户，2010 年的安装基础。2020 年的移动互联网单位是智能手机和平板电脑的安装基础。云和数据中心资本支出包括截至 2022 年结束的十年内的谷歌、亚马逊、微软、 Meta 、阿里巴巴、苹果、 IBM 、甲骨文、腾讯和百度。“ 数十亿台设备 ” 指的是可能使用人工智能技术的潜在设备和用户基础；这包括智能手机、物联网设备、机器人等。来源：Weiss 等人。“ 人工智能指数：绘制 4000 亿美元企业影响 ” 通过 Morgan Stanley (10/23)

AI Technology Compounding =

Numbers Behind The Momentum

AI 技术复合 =

Numbers Behind The Momentum

BOND 13

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

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)

260% 年增长率，跨越十五年 … 用于训练 AI 模型的数据导致了 …

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，包括在公认基准上的最先进改进，>1K 引用，具有历史意义，且使用广泛）。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）是衡量基本处理能力的计算基本单位，表示涉及小数的单次算术运算。在人工智能中，总 FLOPs 通常用于估算训练或运行模型的计算成本。注意：仅显示语言模型（每 Epoch AI，包括在公认基准上的最先进改进，>1K 引用，具有历史意义，使用广泛）。来源：Epoch AI (5/25)

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

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)

… 九年内增长 200%… 从更好的算法中获得的计算收益导致 …

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

注意：估算有多少进展来自更大的模型，而不是更智能的算法，基于没有改进的情况下，达到顶级性能所需的计算能力。来源：Epoch AI (3/24)

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

… 六年增长 150%…… 性能提升来自更优秀的 AI 超级计算机 ……

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

Source: Epoch AI (4/25)

领先 AI 超级计算机性能 (FLOP/s) – 2019-2025，据 Epoch AI

Source: Epoch AI (4/25)

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

Number of New Large-Scale AI Models (Larger than 10^{23} 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^{23} FLOPs or greater, per Epoch AI.
Source: Epoch AI (5/25)

… 四年内增长 167% 在 … 强大 AI 模型的数量

Number of New Large-Scale AI Models (Larger than 10^{23} FLOP*) – 2017-2024, per Epoch AI

* 截至 4/25，'大型 - 规模 AI 模型' 通常定义为每轮训练计算量达到 10_{23} FLOPs 或更高的模型，根据 Epoch AI。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 & 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 & The Information

注意：4/25 的用户数量估计来自 OpenAI CEO Sam Altman 4/11/25 在 TED 演讲中的披露。收入数据是基于 OpenAI 披露的数据。来源：OpenAI 披露的数据（截至 4/25），The Information（4/25）（链接，链接，链接 & 链接）

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

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

达到 365B 年度搜索量 =ChatGPT 比谷歌快 5.5 倍

年度搜索量（B）自谷歌和 ChatGPT 正式推出以来 – 1998-2025，根据谷歌和 OpenAI

注意：虚线条表示谷歌未披露年度搜索量的年份。来源：谷歌公开披露，OpenAI (12/24)。ChatGPT 数据为公司披露的 ~1B 每日查询估计值

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，利用新兴的互联网接入，谷歌开始致力于 ‘组织全球信息，使其普遍可访问且有用。’

近三十年后 – 在人类经历的最快速变化之后 –

大量信息确实已经数字化 / 可访问 / 有用。

我们移动信息的方式正在发生巨
access and ch faster…

…AI 是一个复合器 – 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快速采用。

Knowledge Distribution Evolution =

Over ~Six Centuries

知识传播演化 =

在 ~ 六个世纪里

BOND 22

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

Source: Wikimedia Commons

…Knowledge Distribution – 1993-2021 = Active + Digital Delivery…

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

…Knowledge Distribution – 1993-2021 = Active + Digital Delivery…

互联网 – 公共发布 1993*

* 互联网被广泛认为是在 1993 年随着万维网（WWW）进入公共领域而 ' 公开发布 ' 的，这允许用户创建网站；然而，蒂姆 · 伯纳斯 - 李在 1989 年发明了万维网，根据 CERN 的说法。来源：谷歌，美国国防部，CERN

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

Generative AI – Public Launch of ChatGPT 2022*

ChatGPT 4o

Share

Great, I'll gather quantitative global data on how generative AI is impacting knowledge generation and distribution across all major sectors—including education, research, media, and more.

I'll let you know as soon as the findings are ready.

Research completed in 5m · 25 sources · 116 searches

Global Impact of Generative AI on Knowledge Generation and Distribution

Global Impact of Generative AI on Knowledge Generation & Distribution

Generative AI's rapid advancement since 2022 has led to an explosion in AI-created content and widespread adoption across sectors. Below we compile quantitative global data (2023–2025) illustrating how generative AI is transforming knowledge **generation** (content creation) and **distribution** (sharing and access) in academia, media, education, enterprise, and beyond.

Surge in AI-Generated Content Creation

- **Academic Publications:** AI is now contributing to scientific literature. An analysis found that *1% of all scientific articles published in 2023 showed signs of generative AI involvement* scientificamerican.com . This amounts to an estimated **60,000 research papers** potentially written (in part) by large language models scientificamerican.com . Top journals have responded – by late 2023, 87 of the 100 leading journals had issued guidance on using AI tools in writing macmillan.com – reflecting how commonplace AI-assisted writing has become in research.
- **News and Media Content:** Generative AI is also **writing the news**. By mid-2024, about **6.96% of global news articles were AI-generated** – roughly **60,000 AI-written news articles published per day** newscatcherapi.com . Misinformation trackers have identified **1,271 news and information websites across 16 languages that rely on AI-generated content with minimal human oversight** (as of May 2025) newsguardtech.com , a sharp rise from 467 such sites in 2023 titleformat.com . Even reputable outlets use AI for routine reporting: the Associated Press, for example, was automatically generating **40,000 of its 730,000 news stories (5.5%) via AI** by mid-2023 titleformat.com . This automation has enabled greater volume and speed in news dissemination.

**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

… 知识传播 – 2022+ = 活跃 + 数字 + 生成式交付

生成式 AI – ChatGPT 2022* 的公开发布

ChatGPT 4o

Share

Great, I'll gather quantitative global data on how generative AI is impacting knowledge generation and distribution across all major sectors—including education, research, media, and more.

I'll let you know as soon as the findings are ready.

Research completed in 5m · 25 sources · 116 searches

Global Impact of Generative AI on Knowledge Generation and Distribution

Global Impact of Generative AI on Knowledge Generation & Distribution

Generative AI's rapid advancement since 2022 has led to an explosion in AI-created content and widespread adoption across sectors. Below we compile quantitative global data (2023–2025) illustrating how generative AI is transforming knowledge **generation** (content creation) and **distribution** (sharing and access) in academia, media, education, enterprise, and beyond.

Surge in AI-Generated Content Creation

- **Academic Publications:** AI is now contributing to scientific literature. An analysis found that *1% of all scientific articles published in 2023 showed signs of generative AI involvement* scientificamerican.com . This amounts to an estimated **60,000 research papers** potentially written (in part) by large language models scientificamerican.com . Top journals have responded – by late 2023, 87 of the 100 leading journals had issued guidance on using AI tools in writing macmillan.com – reflecting how commonplace AI-assisted writing has become in research.
- **News and Media Content:** Generative AI is also **writing the news**. By mid-2024, about **6.96% of global news articles were AI-generated** – roughly **60,000 AI-written news articles published per day** newscatcherapi.com . Misinformation trackers have identified **1,271 news and information websites across 16 languages that rely on AI-generated content with minimal human oversight** (as of May 2025) newsguardtech.com , a sharp rise from 467 such sites in 2023 titleformat.com . Even reputable outlets use AI for routine reporting: the Associated Press, for example, was automatically generating **40,000 of its 730,000 news stories (5.5%) via AI** by mid-2023 titleformat.com . This automation has enabled greater volume and speed in news dissemination.

* 我们将 2022 年 11 月 ChatGPT 的公开发布定义为我们认为 AI 的 "*Moment*" 的生成式 AI 的公开发布。ChatGPT 创纪录地实现了最快用户增长（5 天获得 100 万用户）。生成式 AI 能够创建内容的 AI – 文本、图像、音频或代码 – 基于学习到的模式。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；智慧在于它们的简化。

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

AI =

Many Years Before Lift-Off

AI =

起飞前许多年

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

10/50: Alan Turing creates his Turing Test to measure computer intelligence, positing that computers could think like humans

6/56: Stanford computer scientist John McCarthy convenes the Dartmouth Conference on 'Artificial Intelligence,' a term he coined

1/62: Arthur Samuel, an IBM computer scientist, creates a self-learning program that proves capable of defeating a top USA checkers champion

1/66: Stanford researchers deploy Shakey, the first general-purpose mobile robot that can reason about its own actions

AI 'Winter[1]'
(1967-1996)

5/97: Deep Blue, IBM's chess-playing computer, defeats Garry Kasparov, the world chess champion at the time

9/02: Roomba, the first mass-produced autonomous robotic vacuum cleaner that can navigate homes, is launched

10/05: A Stanford team build a driverless car named Stanley; it completes a 132-mile course, winning the DARPA Grand Challenge

4/10: Apple acquires Siri voice assistant & integrates it into iPhone 4S model one year later

6/14: Eugene Goostman, a chatbot, passes the Turing Test, with 1/3 of judges believing that Eugene is human

6/18: OpenAI releases GPT-1, the first of their large language models

6/20: OpenAI releases GPT-3, an AI tool for automated conversations; Microsoft exclusively licenses the model

11/22: OpenAI releases ChatGPT to the public

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 里程碑时间线 – 1950-2022, per Stanford University…

10/50: Alan Turing creates his Turing Test to measure computer intelligence, positing that computers could think like humans

6/56: Stanford computer scientist John McCarthy convenes the Dartmouth Conference on 'Artificial Intelligence,' a term he coined

1/62: Arthur Samuel, an IBM computer scientist, creates a self-learning program that proves capable of defeating a top USA checkers champion

1/66: Stanford researchers deploy Shakey, the first general-purpose mobile robot that can reason about its own actions

AI 'Winter[1]'
(1967-1996)

5/97: Deep Blue, IBM's chess-playing computer, defeats Garry Kasparov, the world chess champion at the time

9/02: Roomba, the first mass-produced autonomous robotic vacuum cleaner that can navigate homes, is launched

10/05: A Stanford team build a driverless car named Stanley; it completes a 132-mile course, winning the DARPA Grand Challenge

4/10: Apple acquires Siri voice assistant & integrates it into iPhone 4S model one year later

6/14: Eugene Goostman, a chatbot, passes the Turing Test, with 1/3 of judges believing that Eugene is human

6/18: OpenAI releases GPT-1, the first of their large language models

6/20: OpenAI releases GPT-3, an AI tool for automated conversations; Microsoft exclusively licenses the model

11/22: OpenAI releases ChatGPT to the public

1: AI '寒冬' 是斯坦福大学计算机科学系的 Kumagai 教授 Nils J. Nilsson 使用的术语，用于描述人工智能在继续取得概念性进展但无法取得任何重大实际成功期间的时期。这随后导致了对人工智能的兴趣和资金的下降。包含来自斯坦福大学以外的来源的数据。来源：斯坦福大学和斯坦福法学院来源，iRobot，TechCrunch，BBC，OpenAI。数据由 BOND 整合。

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

3/23: OpenAI releases GPT-4, a multimodal* model capable of processing both text & images

3/23: Google releases Bard, its ChatGPT competitor

11/23: 28 countries, including USA, EU members & China, sign Bletchley Declaration on AI Safety

4/24: Meta Platforms releases its open-source** Llama 3 model with 70B parameters

5/24: Google introduces AI overviews to augment its search functions

9/24: Alibaba releases 100 open-source Qwen 2.5 models, with performance in line with Western competitors

1/25: DeepSeek releases its R1 & R1-Zero open-source reasoning models

2/25: OpenAI releases GPT-4.5, Anthropic releases Claude 3.7 Sonnet, & xAI releases Grok 3

3/23: Microsoft Integrates Copilot into its 365 product suite

3/23: Anthropic releases Claude, its AI assistant focused on safety & inter-pretability

3/24: USA Department of Homeland Security unveils its AI Roadmap Strategy

5/24: OpenAI releases GPT-4o, which has full multimodality across audio, visual, & text inputs

7/24: Apple releases Apple Intelligence, an AI system integrated into its devices, for developers

12/24: OpenAI announces o3, its highest-ever performing model

1/25: Alibaba unveils Qwen2.5-Max, which surpasses the performance of other leading models (GPT-4o, Claude 3.5) on some reasoning tests

4/25: ChatGPT reaches 800MM weekly users[1]

*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 里程碑时间线 – 2023-2025, per Stanford University

3/23: OpenAI releases GPT-4, a multimodal* model capable of processing both text & images

3/23: Google releases Bard, its ChatGPT competitor

11/23: 28 countries, including USA, EU members & China, sign Bletchley Declaration on AI Safety

4/24: Meta Platforms releases its open-source** Llama 3 model with 70B parameters

5/24: Google introduces AI overviews to augment its search functions

9/24: Alibaba releases 100 open-source Qwen 2.5 models, with performance in line with Western competitors

1/25: DeepSeek releases its R1 & R1-Zero open-source reasoning models

2/25: OpenAI releases GPT-4.5, Anthropic releases Claude 3.7 Sonnet, & xAI releases Grok 3

3/23: Microsoft Integrates Copilot into its 365 product suite

3/23: Anthropic releases Claude, its AI assistant focused on safety & inter-pretability

3/24: USA Department of Homeland Security unveils its AI Roadmap Strategy

5/24: OpenAI releases GPT-4o, which has full multimodality across audio, visual, & text inputs

7/24: Apple releases Apple Intelligence, an AI system integrated into its devices, for developers

12/24: OpenAI announces o3, its highest-ever performing model

1/25: Alibaba unveils Qwen2.5-Max, which surpasses the performance of other leading models (GPT-4o, Claude 3.5) on some reasoning tests

4/25: ChatGPT reaches 800MM weekly users[1]

*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

Top Ten Things AI Can Do Today, per ChatGPT

ChatGPT 根据目前 AI 能做的十件事

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.

Source: ChatGPT (5/15/25)

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.

Source: ChatGPT (5/15/25)

BOND AI = Circa Q2:25 31

AI =

Circa 2030?

AI =

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 =

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 =

前所未有的

BOND 36

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 ... 行业超越学术界 数据 + 计算 + 金融需求上升

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，机器可以从数据中的模式中学习，而无需显式编程。注意：学术界包括由一个或多个机构（包括政府机构）开发的模型。行业 - 学术界合作不包括政府合作，仅捕捉学术机构与行业之间的合作。行业不包括与另一家公司以外的任何实体合作开发的模型。Epoch AI 是一家人工智能索引数据提供者，使用 “ 值得注意的机器学习模型 ” 一词来指定人工智能 / 机器学习生态系统中特别有影响力的模型。Epoch 自 1950 年代以来维护了一个包含 900 个人工智能模型的数据库，根据最先进的技术进步、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，因此某些被一些人认为值得注意的模型可能未被收录。学术模型数量为零并不意味着 2023 年学术界没有生产值得注意的模型，而是 Epoch AI 尚未识别出任何值得注意的模型。此外，学术出版物往往需要更长时间才能获得认可，因为介绍重要架构的高引用率论文可能需要数年才能获得 prominence。中国数据可能因政府限制而受到信息限制。来源：Nestor Maslej 等，《人工智能指数 2025 年度报告》，人工智能指数指导委员会，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：根据这篇标题为"200 万注册开发者，无数突破"的 8/20 博客文章，我们假设 2005 年 NVIDIA 生态系统中的开发者可以忽略不计："要达到 100 万注册开发者花了 13 年，而要达到 200 万则不到两年。"来源：NVIDIA 博客文章、新闻发布、公司概况

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

AI 开发者增长（以 Google 生态系统为代理）=+5 倍至 700 万开发者年增长率

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 = 展开 ··· 网景 IPO 后（1995 年） ... 再次 + ChatGPT 发布后更快（2022 年）

USA Computing-Related* Patents Granted Annually – 1960-2024, per USPTO

*Uses Cooperative Patent Classification (CPC) code G06, which corresponds to computing, calculating or counting patents. Google patents data changes somewhat between each query so numbers are rounded and should be viewed as directionally accurate. Source: USA Patent & Trademark Office (USPTO) via Google Patents (4/25)

USA Computing-Related* Patents Granted Annually – 1960-2024, per USPTO

* 使用合作专利分类（CPC）代码 G06，对应计算、计算或计数专利。谷歌专利数据在每次查询之间有所变化，因此数字会四舍五入，应视为方向性准确。来源：美国专利商标局（USPTO）通过谷歌专利（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

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）基准评估语言模型在数学、法律、医学和历史等 57 个学术和专业学科上的表现。它衡量事实记忆和推理能力，是评估大型语言模型一般知识和解决问题能力的标准。89.8% 是人类表现的一般基准。统计数据显示了每年表现最佳的 AI 模型的平均准确率。来源：Papers With Code via 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 Performance = In Q1:25…
73% of Responses & Rising Mistaken as Human by Testers

% of Testers Who Mistake AI Responses as Human-Generated – 3/25, per Cameron Jones / Benjamin Bergen

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

AI 性能 = 在 2025 年第一季度 …73% 的回复正在上升并被测试者误认为是人类

将 AI 回复误认为是人类生成的测试者百分比 – 3/25，根据卡梅隆 · 琼斯 / 本杰明 · 伯格

注意：图灵测试于 1950 年引入，用于衡量机器模仿人类对话的能力。在本研究中，~500 部分参与者参与了一种三人测试格式，与人类和人工智能进行互动。大多数讨论侧重于情感共鸣和日常话题，而不是事实知识。Eliza 是由麻省理工学院教授约瑟夫 · 韦泽鲍姆于 1960 年代中期开发的，被认为是世界上第一个聊天机器人。在 2025 年 1 月，研究人员使用其原始代码成功复活了 Eliza。来源：卡梅隆 · 琼斯和本杰明伯格，'大型语言模型通过图灵测试'（3/25）通过加州大学圣地亚哥分校

AI Performance = Increasingly Realistic Conversations Simulating Human Behaviors

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

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.**

Source: Cameron Jones and Benjamin Bergen, 'Large Language Models Pass the Turing Test' (3/25) via UC San Diego

AI Performance = Increasingly Realistic Conversations Simulating Human Behaviors

与 GPT-4.5 的图灵测试对话 – 3/25，据卡梅隆 · 琼斯 / 本杰明 · 伯格

测试内容：
图灵测试是由艾伦图灵在 1950 年提出的，用于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答来自人类还是机器，则该机器被认为通过了测试。在这里，参与者必须猜测见证人 A 或见证人 B 是人工智能系统。

结果：
左侧的对话是一个于 3/25 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 将左侧图片 (证人 A) 判定为人类，确定度为 87%，并说 'A 有人的感觉。B 有人工智能模仿的感觉。' A 实际上是人工智能生成的；B 是人类的。

来源：卡梅隆 · 琼斯和本杰明 · 伯格，《大型语言模型通过图灵测试》（3/25）通过加州大学圣地亚哥分校

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- 生成图像：'带向日葵吊坠的女式项链' – 2/22-4/25，每 Midjourney / 金企鹅

模型 v1 (2/22)

模型 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 性能 = 图像生成越来越逼真

AI 生成 vs. 真实图像 – 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…

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

AI Performance = Increasingly Realistic Audio Translation / Generation…

ElevenLabs AI 语音生成器 –1/23-4/25，根据 ElevenLabs & Similarweb

当您创建一个新的配音项目时，配音工作室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频轨道。在生成翻译之前，会隔离并克隆每个说话者的原始声音，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 的数百万用户已生成 1000 年的音频内容，该公司的工具已被《财富》 500 强公司中超过 60% 的员工采用。

- ElevenLabs Press Release, 1/25

注意：由于政府限制，中国数据可能存在信息限制。来源：ElevenLabs （1/24 & 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 – 2025 年 5 月 25 日，据 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，并且你在某个领域是世界专家 …… 但恰好你是印度尼西亚人。如今，存在语言障碍，如果你不懂英语，想要走向世界舞台将非常困难。但有了人工智能，未来或许可能实现，你用母语说话，人工智能能够理解并用实时翻译 ……

…… 这对创造力、知识分享、娱乐将产生什么影响？我认为我们正处于探索这个问题的非常早期的阶段 ……

…… 我们希望 Spotify 成为所有声音的归宿。

- Spotify 联合创始人兼 CEO 丹尼尔 · 艾克 (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 2025 年第一季度数据年化计算。来源：Spotify、'纽约邮报'、'Spotify 内部：CEO 丹尼尔 · 艾克谈人工智能、免费语音与音乐的未来'（2025 年 5 月 2 日）；Spotify 财报发布；eMarketer、'Spotify 在数字音频领域超越苹果和亚马逊'（2025 年 4 月 25 日）

AI Performance = Emerging Applications Accelerating

Emerging AI Applications – 11/24, per Morgan Stanley

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

AI 性能 = 新兴应用加速

新兴 AI 应用 – 11/24，根据摩根士丹利

摩根士丹利，《生成式 AI：我们正在看到哪些领域的采用以及对 2025 年有何重要意义？》（11/24）

AI =

Benefits & Risks

AI =

益处与风险

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

AI 的广泛讨论的益处和风险 – 是许多人的首要考虑 – 引发了应得的兴奋 和 恐惧 , 进一步加剧了对快速变化的不确定性、日益加剧的全球竞争和军事威胁的担忧。

Stuart Russell 和 Peter Norvig 在他们的 1,116- 页经典著作《人工智能：一种现代方法》（链接在此），深入探讨了这些主题，并且他们的观点仍然适用。

重点如下 …

… 益处：简单来说，我们整个文明是人类智慧的产物。如果我们能够获得大幅提升的机器智能，那么 [我们的] 雄心壮志的极限将大幅提高。

人工智能和机器人的潜力可以解放人类免于繁琐重复的工作，并大幅提高商品和服务的生产，这可能预示着一个和平富足的时代。加速科学研究的能力可能导致疾病的治疗方案以及应对气候变化和资源短缺的解决方案。

正如谷歌 DeepMind 的首席执行官 Demis Hassabis 所建议的：“ 首先我们解决人工智能问题，然后利用人工智能解决其他一切问题。” 然而，在我们有机会 “ 解决人工智能问题 ” 之前，我们将因人工智能的滥用而面临风险，无论是有意还是无意的。

其中一些已经显现，而另一些则根据当前趋势似乎可能包括致命性自主武器 …… 监控和说服 …
… 有偏见的决策 …… 对就业的影响 …… 安全 - 关键应用 …… 网络安全 ……

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)

Source: University of Cambridge, Centre for the Future of Intelligence

成功创造人工智能可能是我们文明历史上最大的事件。但它也可能是最后一个 – 除非我们学会如何避免风险。

斯蒂芬 · 霍金，理论物理学家 / 宇宙学家（1942-2018）

Source: University of Cambridge, Centre for the Future of Intelligence

Outline

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

Outline

1. 似乎变化比以往任何时候都快？是的，就是这样
2. AI 用户 + 使用 + 资本支出增长 = 前所未有的
3. AI 模型计算成本高 / 上升 + 每令牌推理成本下降 = 性能趋同 + 开发者使用量上升
4. AI 使用 + 成本 + 损失增长 = 前所未有的
5. 人工智能变现威胁 = 竞争加剧 + 开源势头 + 中国崛起
6. AI 与物理世界加速 = 快速 + 数据驱动
7. 由 AI 驱动，从一开始就加速全球互联网用户增长 = 前所未见的增长
8. AI 与工作进化 = 真实的 + 快速的

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

Unprecedented

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

前所未有的

BOND 53

Consumer / User AI Adoption =

Unprecedented

Consumer / User AI Adoption =

前所未有的

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

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

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

Internet vs. ChatGPT Users – Percent Outside North America (1990-2025), Per ITU & Sensor Tower

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)

AI 全球采用（ChatGPT 作为基础指标） = 以前没见过这样的全球传播

Internet vs. ChatGPT Users – Percent Outside North America (1990-2025), Per ITU & Sensor Tower

注意：互联网的年份 = *1990*； 年份 *33 = 2022*。ChatGPT 应用的年份 = *5*/23； 年份 3 的 ChatGPT 应用 = *5*/25。显示 ChatGPT 应用的月活跃用户（MAU）。请注意，截至 5 月 25 日，ChatGPT 在中国、俄罗斯和某些其他国家不可用。由于政府限制，中国的数据可能受到信息限制。仅包括 Android、 iPhone 和 iPad 用户。数据可能低估了真正的 ChatGPT 用户基数（例如，桌面或移动网页用户）。根据联合国定义的地区。数据显示为该年总用户数的百分比 – 请注意，由于 ChatGPT 的年份 3 尚未结束，未来几个月百分比可能会变化。仅针对独立 ChatGPT 应用。根据国际电信联盟，某些年份的国家级数据可能缺失。来源：联合国 / 国际电信联盟（3 月 25 日）、 Sensor Tower （5 月 25 日）

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

Years to Reach 100MM Users – 2000-2023

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

AI 用户采用（以 ChatGPT 作为代理） = 实质上更快与互联网可比对象 …

达到 1 亿用户所需年份 – 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

Days to Reach 1MM Customers / Users – 1908-2022

*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

…AI 用户采用（ChatGPT 作为代理） = 实质上更快 + 更便宜与其他基础技术产品相比

达到 100 万客户 / 用户所需天数 – 1908-2022

*ChatGPT 公开发布 = 首次向公众作为免费研究预览发布（11/22）。注意：根据福特公司的说法，T 型车可以以 260 至 850 美元的价格出售。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 日的消费者销售启动日期，当时 TiVo 为其 14 小时盒式录像机收取 499 美元。我们不计算 TiVo 订阅成本。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, 美国速度博物馆 , 世界银行 , 福特公司 , Gizmodo, 苹果 , 大英百科全书 , 圣路易斯联邦储备银行 , 维基媒体共享 , UBS

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

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

AI 用户采用 – 达到 50% 家庭渗透率 = 每个周期在 ~ 不到一半的时间内加速 …AI 遵循该模式

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

注意：AI 时代 3 年意味着达到 50% 美国家庭采用率的时间与前一个周期相比同样减半。Source: 摩根士丹利，《谷歌和 Meta：AI 与基本面第二半场辩论》（7/23），我们的世界数据，根据 MS 的其他网络来源

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 Computing Ecosystem – 2021-2025, per NVIDIA

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

NVIDIA AI 生态系统在开发者 / 创业公司 / 应用方面增长超过四年 =>100%

NVIDIA 计算生态系统 – 2021-2025, per NVIDIA

注意：GPU = 图形处理器。来源：NVIDIA (2021 & 2025)

Tech Incumbent AI Adoption =

Top Priority

Tech Incumbent AI Adoption =

Top Priority

BOND 62

Tech Incumbent AI Focus = Talking-the-Talk…

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

技术巨头 AI 关注 = 谈谈 …

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

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

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

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

Source: Amazon (4/10/25), Google (4/9/25), Techradar

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

生成式 AI 将重塑我们几乎所知的所有客户体验，并使我们能够实现关于我们仅凭幻想才能想到的全新体验。正在部署的早期 AI 工作负载专注于提高生产力和避免成本 ……

… 越来越多地，你会看到 AI 改变编码、搜索、购物、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 – 所有领域。

- 亚马逊 CEO 安迪 · 贾西在 2024 年亚马逊股东信 –4/25

改善生活和重新构想事物的机会是为什么谷歌已经在人工智能领域投资了十多年 ……

… 我们将其视为我们推进组织世界信息、使其普遍可访问和有用的使命的最重要方式 …

…AI 带来的机遇是最大的。

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

Source: Amazon (4/10/25), Google (4/9/25), Techradar

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

duolingo

AI Going Full-Circle:

DeepMind's AlphaGo (2014) started with humans training machines…Duolingo Chess now has machines training humans…

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

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

duolingo

AI Going Full-Circle:

DeepMind 的 AlphaGo (2014) 最初是让人类训练机器 ……Duolingo Chess 现在是让机器训练人类 ……

有三个地方，*[GenAI 正在]*… 帮助我们：数据创建 … 创建以前不可能的新功能 … 公司内到处都是效率 …

… 我应该提一下关于 *[新 Duolingo 课程在]* 棋类方面的惊人之处是它真正始于一个由两个人组成的团队，其中没有人会编程 … 他们基本上制作了原型，并通过仅使用 AI 完成了整个棋类课程。他们中也没有人会下棋。

- D Duolingo 联合创始人兼 CEO 路易斯 · 冯 · 阿恩 @ Q1:25 财报电话会议 **–** 5/25

Grok 的 AI 正变得非常出色 …… 重要的是要为 AI 编程设定良好的价值观，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地追求真理的 AI。

- xAI 创始人兼 CEO 埃隆 · 马斯克 **–**5/25

Note: 在 3/28/25，埃隆 · 马斯克宣布 xAI 通过全股票交易收购了 X。该交易使 xAI 估值 80B 美元，X 估值 33B 美元（扣除 12B 美元的债务后为 45B 美元）。来源：Duolingo (5/1/25)、 DeepMind 、埃隆 · 马斯克 (5/2/25)、福克斯新闻

…Tech Incumbent AI Focus = Talking-the-Talk

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

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

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

… 技术巨头 AI 关注 = 说空话

我们认为 AI 是一种人类加速工具，将使个人能够做更多我相信长期来看，我们将看到 … 人们与 …

the AI they use as the overall output of that person.

- Roblox 联合创始人、总裁、 CEO 兼董事会主席 David Baszucki @ Q1:25 财报电话会议 – 5/25

我向你保证，十年后，你会回想起并意识到 AI 已经融入了所有事物。事实上，我们需要无处不在的 AI。

而且每个地区、每个行业、每个国家、每家公司，都需要 AI。AI *[* 是 *]* 现在已成为基础设施。而这个基础设施，就像互联网，就像电力，需要工厂 ……。

… 而这些 AI 数据中心，如果你愿意这么说的话，被不恰当地描述了。实际上，它们是，AI 工厂。你输入能量，它就能产出极其有价值的东西。

- 英伟达联合创始人兼 CEO 黄仁勋 @
COMPUTEX 2025 – 5/25

来源：Roblox (5/1/25), NVIDIA (5/18/25)

‘Traditional’ Enterprise AI Adoption =

Rising Priority

‘传统’ 企业人工智能采用 =

日益重要的优先级

BOND 67

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

季度盈利电话会议中提及 '人工智能' —— 标普 500 公司（2015-2025），根据高盛研究

Source: Goldman Sachs Global In投资研究，'标普米色手册：2024 年第四季度电话会议的 3 个主题：关税、强劲的美元和人工智能'（2/25)

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

GenAI Improvements Targeted for Global Enterprises over Next 2 Years – 2024, per Morgan Stanley

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)

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

未来两年针对全球企业的 GenAI 改进 – 2024 年，根据摩根士丹利

注意：调查于 5 月 24 日进行，样本量 N=427。美国公司占 = 43%，日本 15%，英国 14%，法国 14%，德国 14%。行业分布：18% 技术，18% 金融服务，17% 医疗保健，17% 制造业，15% 工业，15% 消费品。收入分布：13% 5000 万 -7500 万美元，25%7510 万 -1 亿美元，36%1 亿 -5 亿美元，10%5 亿 -10 亿美元，8%10 亿 -15 亿美元，3%15 亿 -$20B，5% $20B+。'收入 - 导向' 和 '成本 - 导向' 分类依据 BOND，而非摩根士丹利。来源：AlphaWise，摩根士丹利，《量化 AI 机遇》（24 年 12 月）

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

Global Chief Marketing Officer (CMO) GenAI Adoption Survey – 2024, per Morgan Stanley

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)

企业 AI 焦点 – 全球 CMOs =75% 使用 / 测试 AI 工具

Global Chief Marketing Officer (CMO) GenAI Adoption Survey – 2024, per Morgan Stanley

注意：调查问题询问全球营销高管在营销活动中使用生成式 AI 的程度。调查于 7/24 进行，N = 300 全球拥有 500+ 名员工的公司营销高管。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、 & 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)

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

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

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)

企业人工智能采用 = 日益重要 … 美国银行 – Erica 虚拟助手 (6/18)

美国银行 Erica 虚拟助手 – 6/18-2/25，根据美国银行

注意：埃里卡是一款集成在美国银行移动应用中的对话式人工智能，通过提供实时洞察、交易搜索、账单提醒和预算辅助等功能，帮助客户管理财务。它已处理数十亿次互动，并为超过 4000 万客户提供全天候的数字金融管家服务。

Erica 兼具个人管家和任务控制中心的作用，为我们的客户服务。

我们的数据科学团队自发布以来已对 Erica 的性能进行了超过 50,000 次更新 —— 调整、扩展和优化自然语言理解能力，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引人注目的里程碑，但这只是 Erica 的开始。

- 美国银行数字业务负责人妮基 · 卡茨，4/24

Note: We 假设从 6/18 的 Erica 发布开始，用户数为零。试点用户除外。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)

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

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

JP Morgan Estimated Value from AI / ML

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)

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

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

我们对通过 AI 获得的效率提升充满期待

… 某些关键用户子集告诉我们他们每周获得了数小时的生产力，并且几乎可以肯定，时间节省来自于价值较低的任务 …

… 我们在人工智能领域是早期参与者。但我们仍处于这一旅程的早期阶段。

- 摩根大通首席财务官杰里米 · 巴 NUM，5/25

JP Morgan Estimated Value from AI / ML

注意：根据摩根大通，上标 '2' 表示 ' 价值被描述为收入增加、费用减少或避免成本 —— 多数情况下以相对于先前的分析方法提升的幅度衡量，其余部分以相对于随机基线或保留控制组的幅度衡量。' 我们标示 2020 年为摩根大通人工智能现代化（2020 年致股东信：' 我们已经广泛使用人工智能，在欺诈和风险、营销、潜在客户开发、创意生成、运营、交易和其他领域取得了相当大的成功 —— 但我们仍处于这一旅程的早期阶段 ' ）。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

KAISER PERMANENTE | About

News | Health & Wellness | Who We Are | Commitments & Impact

Home | News | Press release archive | Kaiser Permanente improves member experience with AI-enabled clinical technology

August 14, 2024

Kaiser Permanente improves member experience with AI-enabled clinical technology

Assisted clinical documentation tool from Abridge helps doctors reduce time spent on administrative tasks, allowing them to be more present with patients.

PRESS RELEASE
Contact: Vincent Staupe
vincent.p.staupe@kp.org

OAKLAND, Calif. — Kaiser Permanente, the nation's largest nonprofit, integrated health care provider, today announced that a new assisted clinical documentation tool is now available to doctors and other clinicians at Kaiser Permanente's 40 hospitals and more than 600 medical offices in eight states and the District of Columbia. The tool, powered by ambient listening technology from Abridge, supports doctors and other clinicians with securely capturing clinical notes during visits with patients and helps them remain focused on talking with patients rather than on documentation or administrative tasks.

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

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

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)

企业 AI 采用 = 日益优先 …Kaiser Permanente – 多模态环境 AI 笔记本 (10/23)

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

KAISER PERMANENTE | About

News | Health & Wellness | Who We Are | Commitments & Impact

Home | News | Press release archive | Kaiser Permanente improves member experience with AI-enabled clinical technology

August 14, 2024

Kaiser Permanente improves member experience with AI-enabled clinical technology

Assisted clinical documentation tool from Abridge helps doctors reduce time spent on administrative tasks, allowing them to be more present with patients.

PRESS RELEASE
Contact: Vincent Staupe
vincent.p.staupe@kp.org

OAKLAND, Calif. — Kaiser Permanente, the nation's largest nonprofit, integrated health care provider, today announced that a new assisted clinical documentation tool is now available to doctors and other clinicians at Kaiser Permanente's 40 hospitals and more than 600 medical offices in eight states and the District of Columbia. The tool, powered by ambient listening technology from Abridge, supports doctors and other clinicians with securely capturing clinical notes during visits with patients and helps them remain focused on talking with patients rather than on documentation or administrative tasks.

环境式人工智能（AI）笔记本，它利用机器学习应用于对话以实时实现类似笔记的功能，[有] 巨大潜力以减轻文档负担、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实时转录会话，但不会保留音频记录。为应对日益紧迫和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式 AI 技术，以增强他们在不同环境和专业领域的临床能力。

- 新英格兰医学杂志催化剂研究报告，2/24

独特的凯泽医疗保健医师使用 AI Scribe & 累计 Scribe 访问次数

来源：Tierney, Aaron A. 等人，《环境式人工智能笔记本以减轻临床文档负担》(3/24) & Tierney, Aaron A. 等人，《环境式人工智能笔记本：一年后的学习与超过 250 万次使用》(3/25) 通过 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)

Enterprise AI Adoption = Rising Priority… 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

Yum! Restaurants Using at Least One Byte by Yum! Product

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)

企业 AI 采用 = 上升优先级 …Yum! Brands – Byte by Yum! (2/25)

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

Byte 是 Yum! Brands 的 AI 驱动餐厅管理平台，旨在通过自动化库存跟踪、排班和食品准备警报等重复性任务来优化店铺运营。它利用机器学习来改善餐厅层面的决策，提高效率，减少浪费，并支持员工生产力。

由人工智能支持，Byte by Yum! 为特许经营者提供领先的科技能力，并通过 Yum! 的规模实现优势经济。

Byte by Yum! 平台包括在线和移动应用订餐、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! Press Release, 2/25

Yum! 餐厅使用至少一款 Byte by Yum! 产品

注意：Yum! Brands 名称包括肯德基、塔可钟、必胜客和 The Habit。Byte by Yum! 于 2/25 正式推出。虽然底层技术此前已在 Yum! 的餐厅组合中使用，但 Byte by Yum! 产品套件尚未正式推出；因此，我们在 2/24. 来源：Yum!，'推出 Byte byYum!，一个由人工智能驱动的餐厅技术平台，为全球客户和团队成员提供体验'(2/25)

Education / Government / Research AI Adoption =

Rising Priority

Education / Government / Research AI Adoption =

优先级提升

Education & Government = Increasingly Announcing AI Integrations

教育 & 政府 = 越来越多地宣布 AI 集成

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

New team of technologists creating artificial intelligence (AI) tools

Oxford Partnership – 3/25

5-Year Partnership on Research & AI Literacy

NextGenAI – 3/25

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

ChatGPT Gov – 1/25

ChatGPT tailored for USA federal agencies

USA National Laboratories – 1/25

Partnering on Nuclear, Cybersecurity, & Scientific Breakthroughs

Source: Arizona State University (8/23), Oxford University (3/25), University of Michigan (3/25), Launch Consulting (1/25) via AI Advantage Daily News, NPR (1/25)

亚利桑那州立大学的 "AI 加速" – 8/23

新的一批技术专家正在创建人工智能（AI）工具

牛津合作 – 3/25

5-Year Partnership on Research & AI Literacy

NextGenAI – 3/25

$50MM 联合体，包含 15 所研究型大学（MIT、哈佛大学、加州理工学院等）

ChatGPT Gov – 1/25

ChatGPT tailored for USA federal agencies

USA National Laboratories – 1/25

在核能、网络安全及科学突破方面进行合作

Source: Arizona State University (8/23), 牛津大学 (3/25), 密歇根大学 (3/25), Launch Consulting (1/25) via AI Advantage Daily News, NPR (1/25)

Government = Increasingly Adopting Sovereign AI Policies

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

Source: NVIDIA (2/25 & 5/25)

政府 = 越来越多地采用主权人工智能政策

NVIDIA 主权人工智能合作伙伴 – 2/25, Per NVIDIA

各国正在投资人工智能基础设施，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黄仁勋，5/25

Source: NVIDIA (2/25 & 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

AI-Enabled Medical Devices Approved

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

新 AI 赋能医疗器械获美国食品药品监督管理局批准 –1995-2023，根据斯坦福 HAI 与美国 FDA

获批的 AI 医疗设备

New USA FDA AI Policy (5/25)

在 USA FDA 的历史性首次中，[美国食品药品监督管理局], FDA 局长马丁 ·A· 马卡里，医学博士、公共卫生硕士，今天宣布了一个积极的时间表，计划在 2025 年 6 月 30 日前，在 FDA 所有中心内部全面推广使用人工智能（AI）……

… 为反映这项工作的紧迫性，Makary 博士已指示所有 FDA 中心立即开始部署，目标是在 6 月底前实现全面整合。

- USA FDA 新闻稿，5/25

注意：2021 财年、2022 财年及 2023 财年美国政府的预算数据为实际数据。2024 财年的数据已通过立法但尚未实际执行，2025 财年的数据为申请数据。NIH 在总预算中的份额为申请数据。来源：Nestor Maslej 等，《AI 指数 2025 年度报告》，AI 指数指导委员会，斯坦福 HAI （4/25）；美国食品药品监督管理局，《FDA 宣布完成首次 AI 辅助科学审查试点和积极机构范围 AI 推广时间表》（5/25）；NITRD.gov （5/25）

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)

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

使用 Cradle 的制药公司通过使用我们的 GenAI 平台设计生物制剂，在临床前研究和开发中实现了 1.5 倍到 12 倍的加速。

- Stef van Grieken, Cradle 联合创始人兼 CEO，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关卡，并正式移交给开发组织进行与开始人体临床试验相关的工作。数据收集时间范围为 2021-2024 年。来源：Cradle, Insilico Medicine via BioPharmaTrend，'Insilico Medicine 发布其 AI- 设计疗法的基准' （2/25）

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

Unprecedented

AI 用户 + 使用 + 资本支出增长 =

前所未有的

BOND 80

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

% of USA Adults Who Say They Have Ever Used ChatGPT – 7/23 per Pew & 1/25 per 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 – ChatGPT = 在美国各年龄段中迅速增长，根据皮尤研究中心和伊隆大学的数据

% of USA Adults Who Say They Have Ever Used ChatGPT – 7/23 per Pew & 1/25 per Elon University

注意：7/23 的数据来自皮尤研究中心关于 ChatGPT 使用的调查，n=10,133 名美国成年人。未给出答案的人不被显示。1/25 的数据来自伊隆大学关于使用任何 AI 模型的研究，n=500 名美国成年人。这些数字基于整体 AI 工具使用率调整，考虑到使用任何 AI 工具的受访者中 ChatGPT 的平均使用率为 72%。实际的 ChatGPT 渗透率可能因群体而异。请注意，这张图表汇总了来自不同调查提供者的数据，因此可能无法直接比较。来源：皮尤研究中心（2024 年 3 月 26 日），伊隆大学（2025 年 3 月 12 日发布），山姆·奥尔特曼（2025 年 5 月 12 日）通过《福布斯》

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

Minutes per Day that USA Active Users Spend on 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 参与度（以 ChatGPT App 作为代理） =+202% 在超过二十一个月的每日使用时间中增长 …

美国活跃用户每天在 ChatGPT App 上花费的分钟数 – 7/23-4/25，来自 Sensor Tower

注意：数据代表美国 App Store 和 Google Play Store 的月活跃用户。仅包含 ChatGPT 独立应用的数据。截至 5/25，ChatGPT 应用在中国、俄罗斯和部分其他国家不可用。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 参与度（作为代理的 ChatGPT 应用程序） =+106% 增长的会话次数 & +47% 增长的持续时间，跨越二十一个月

ChatGPT 应用程序的 USA 平均会话时长（分钟）和每日用户会话次数 –7/23-4/25，根据 Sensor Tower 数据

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

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

Consumer ChatGPT & Google Search Global Desktop User Retention Rates (1/23-4/25), 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 留存率（以 ChatGPT 作为代理） =80% vs. 58% 在二十七个月内，根据 YipitData

Consumer ChatGPT & Google Search Global Desktop User Retention Rates (1/23-4/25), per YipitData

注意：留存率 = 指从上一周留存至当前周的用户百分比。数据测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 和网络浏览器扩展）中收集。用户必须连续两个月参与面板才能被纳入。面板在全球范围内具有代表性，但中国的数据可能因政府限制而存在信息限制。排除了 2023 年 12 月 24 日、 12 月 31 日、 2024 年 12 月 22 日、 12 月 29 日、 2025 年 1 月 5 日的异常情况，可能由于假期导致企业使用减少。来源：YipitData （ 5 月 25 日）

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 聊天机器人 @ 工作 告诉 =>72% 更快 / 更好地完成事情

% 的美国在职成年人使用 AI 聊天机器人，他们说工具在 _____ **在 …… 方面很有帮助** – 10/24，根据皮尤的数据

注意：对 273 名美国成年人进行调查，他们兼职或全职工作，并且只有一份工作或有多份工作但认为其中一份是他们的主要工作。来源：皮尤研究中心（10/24）

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

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)

AI 聊天机器人 @ 学校使用（以 ChatGPT 作为代理） = 倾向于研究 / 问题解决 / 学习 / 建议

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 名美国大学和研究生，涵盖 STEM 和非 STEM 学科；仅使用 18-24 岁人群的回答。样本包括 AI 用户和非用户，但排除了"AI 拒绝者"—— 定义为在未来 12 个月内对采用 AI 几乎没有兴趣的非用户。来源：OpenAI,' 构建一个 AI- 准备好的劳动力：对美国大学生 ChatGPT 采用的看法 ' （2/25）

AI Usage Expansion – Deep Research = Automating Specialized Knowledge Work

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

Google Gemini Deep Research

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

OpenAI ChatGPT Deep Research

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

xAI Grok DeepSearch

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

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

AI 应用扩展 – 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

Google Gemini 深度研究

使用 Deep Research，Gemini 中的一个智能代理功能，可以自动为您浏览多达数百个网站，思考其发现，并创建有见地的多页报告，您可以将其转化为引人入胜的播客式对话 …

… 这是迈向更智能的 AI 的一步，它可以超越简单的问答，成为真正的协作伙伴。

- Google 深度研究概述，发布于 12/24

OpenAI ChatGPT 深度研究

Today we're launching deep research in ChatGPT, a new agentic capability that conducts multi-step research on the internet for complex tasks.
它用几分钟完成了人类需要花费数小时才能完成的工作 …

… 深度研究标志着我们实现更广泛目标的重要一步，即开发通用人工智能（AGI），这是我们长期以来一直设想的能够产出新型科学研究成果的能力。

- OpenAI 深度研究新闻稿，2/25

xAI Grok DeepSearch

要理解宇宙，我们必须让 Grok 与世界接口 ……

…… 作为实现这一愿景的第一步，我们正在推出 DeepSearch– 我们的第一个智能体。它是一个闪电般的 AI 智能体，旨在无情地寻求人类知识库中的真相。DeepSearch 旨在综合关键信息，推理冲突的事实和观点，并从复杂性中提炼清晰度。

- xAI Grok 3 Beta 新闻稿，2/25

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.

传统的聊天机器人被设计为响应用户提示，通常在僵化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 – 但始终在反应性、有限的框架内。

AI 代理代表了一个向前跨越的步骤。这些是智能的长时间运行的过程，可以进行推理、行动，并代表用户完成多 **步任务。它们不只是回答问题** – 它们执行：预订会议、提交报告、登录工具，或跨平台协调工作流程，通常使用自然语言作为它们的命令层。

这种转变反映了技术中更广泛的历史模式。正如 2000 年代初静态网站让位于动态网络应用程序 – 工具如 Gmail 和 Google Maps 将互联网从一个页面集合转变为一系列实用工具 –AI 代理正在将对话界面转变为功能基础设施。

而早期的助手需要清晰的输入并产生狭窄的输出，代理承诺以目标、自主性和一定的护栏来运行。它们承诺解释意图、管理内存，并协调跨 **应用** 完成实际工作。这与其说是响应，不如说是完成。

虽然这些智能体的开发尚处于早期阶段，但其影响才刚刚开始显现。智能体可能会重塑用户与数字系统的交互方式 – 从客户支持和入职到研究、日程安排和内部运营。

企业正在引领潮流；它们不仅在试验智能体，还在部署它们，投资于框架并在自主执行周围构建生态系统。曾经只是一个消息界面正在变成一个操作层。

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

全球谷歌搜索 'AI Agent' （K） – 1/24-5/25, per 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
Salesforce Agentforce *(10/24 = General Release)*	• Automated customer support • Case resolution • Lead qualification • Order tracking
Anthropic Claude 3.5 Computer Use *(10/24 = Research Preview Release)*	• Control computer screen directly to perform tasks like pulling data from websites, making online purchases, etc.
OpenAI Operator *(1/25 = Research Preview Release)*	• Control computer screen directly to perform tasks like pulling data from websites, making online purchases, etc.
Amazon Nova Act *(3/25 = Research Preview Release)*	• Home automation • Information collection • Purchasing • Scheduling

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

AI Agent Deployments = AI Incumbent Product Launches Accelerating

AI Incumbent Agent Launches

Agent Released	Select Capabilities
Salesforce Agentforce *(10/24 = General Release)*	• 自动化客户支持 案例解决 • • 潜在客户资格确认 • 订单跟踪
Anthropic Claude 3.5 Computer Use *(10/24 = Research Preview Release)*	• 直接控制电脑屏幕执行任务，如从网站获取数据、进行在线购买等。
OpenAI Operator *(1/25 = Research Preview Release)*	• 直接控制电脑屏幕执行任务，如从网站获取数据、进行在线购买等。
Amazon Nova Act *(3/25 = Research Preview Release)*	• 家庭自动化 • 信息收集 • Purchasing • Scheduling

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

BOND 92

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，或称 AGI，是指能够执行人类全部智力任务的系统 – 推理、规划、从小数据样本中学习以及在跨领域泛化知识。与当前在特定（尽管广泛）边界内表现优异的人工智能模型不同，AGI 将能够跨学科完全灵活地运作，并在无需重新训练的情况下解决不熟悉的问题。它代表了人工智能发展中的一个重要里程碑 – 这一里程碑建立在近期模型规模、训练数据和计算效率的指数级增长之上。

AGI 的时间表仍不确定，但专家预期近年来已显著提前。OpenAI CEO 萨姆 · 奥特曼在 2025 年 1 月评论道，我们现在有信心知道如何按照传统理解构建 AGI。 这是一种预测，而非定论，但它反映了模型架构、推理 * 效率以及训练规模的进步正在缩短研究与前沿能力之间的距离。更广泛的趋势清晰可见：人工智能发展正以前所未有的速度进行，AGI 正越来越多地被视为一个可触及的阈值，而非一个假设的终点。

如果 / 当实现时，通用人工智能（ AGI ）将重新定义软件（及相关硬件）能做什么。AGI 系统将理解目标、生成计划，并实时自我纠正。它们可以在几乎没有人类监督的情况下驱动研究、工程、教育和物流工作流程 – 使用通用推理处理模糊性和新颖性。这些系统不需要大量重新训练来处理新的问题领域 – 它们将迁移学习并具有上下文，就像人类专家一样。此外，由 AGI 驱动的拟人化机器人将拥有重塑我们物理环境以及我们如何在其中运作的能力。

然而，这些影响值得谨慎看待。AGI 不是一个终点，而是一个能力转变的阶段 – 它如何重塑机构、劳动和决策将取决于伴随它的安全措施和部署框架。生产力提升可能很大，但分布不均。

地缘政治、伦理和经济影响可能会逐渐发展，而不是突然发生。与早期的转变 – 从工业到数字再到算法 – 其全部后果不仅将由技术能做什么决定，还由社会选择如何采用和管理它。

* 推理 = 全训练模型根据用户输入生成预测、答案或内容。这一阶段比训练快得多，也更高效

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

Unprecedented

AI 用户 + 使用 + 资本支出增长 =

前所未有的

BOND 94

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（CapEx）的走向，了解其历史发展有助于我们理解。在过去的二十年里，技术资本支出在数据漫长的发展历程中，在某些时刻向上发展 —— 首先向存储 / 访问发展，然后向分发 / 扩展发展，现在向计算 / 智能发展。

最早的浪潮中，资本支出投入于构建互联网基础设施 —— 庞大的服务器农场、海底光缆和早期数据中心，这些设施使亚马逊、微软、谷歌等公司奠定了云计算的基础。那是第一个阶段：存储它，组织它，服务它。

第二波 —— 仍在进行中 —— 主要关于为数据密集型 AI 工作负载超级增强计算能力，这是云计算的自然演变。超大规模计算 * 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 还是一个研究功能；到 2023 年，它已经成为一项资本支出项目。微软副董事长兼总裁布拉德 · 史密斯在 4 月 25 日的博客文章中很好地表达了这一点：
就像过去电力和其他通用技术一样，AI 和云数据中心代表了工业化下一个阶段。

全球最大的科技公司每年都在花费数百亿美元 – 不仅仅是收集数据，而且是从中学习，用其推理，并实时变现。这仍然关于数据 – 但现在，优势在于那些能够最快地训练它、最深入地个性化它、最广泛地部署它的人。

* 超大规模计算器（大型数据中心运营商）是亚马逊网络服务（AWS）、微软 Azure 、谷歌云平台（GCP）、阿里巴巴云、甲骨文云基础设施（OCI）、IBM 云和腾讯云。

CapEx Spend – Big Technology Companies =

On Rise for Years as
Data Use + Storage Exploded

CapEx Spend – 大型科技公司 =

多年来持续增长，数据使用
+ 存储激增

BOND 96

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

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

CapEx Spend @ 大型六家 * 科技公司 (美国) =+21% 十年年度增长

大型六家 * 美国公共科技公司 CapEx Spend ($B) 与全球数据生成 (泽字节) – 2014-2024, 根据 Capital IQ & Hinrich Foundation

* 注意：美国大型六家科技公司包括苹果、英伟达、微软、 Alphabet / 谷歌、亚马逊 & Meta Platforms / Facebook。仅显示亚马逊的 AWS CapEx & 收入（即不包括亚马逊零售 CapEx ）。AWS CapEx 根据摩根士丹利 – 估算，等于 AWS 净增加的物业及设备减去金融租赁和义务。2024 年全球数据生成数据为估算值。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

技术超大规模企业的资本支出 = 与 … 相呼应 +37% 年云收入增长十年间

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（亿美元） – 2014-2024，根据公司披露与摩根士丹利估计

注意：公司未按同类比较方式报告 " 同类云收入 "，因此数据代表最佳估计值，公司之间可能不完全一致。甲骨文云收入包括云服务与许可支持，以及云许可与在用本地许可。IBM 云收入因报告标准而包括所有 " 基础设施 " 项目。阿里云与腾讯云收入根据摩根士丹利估计。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 Spend – 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 Model Training Dataset Size (Tokens) by Model Release Year – 6/10-5/25, per 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)

AI 模型训练数据集大小 =250% 年增长跨越十五年，每 Epoch AI

AI 模型训练数据集大小（Tokens）按模型发布年份 – 6/10-5/25，每 Epoch AI

注意：在 AI 语言模型中，tokens 代表文本的基本单元（例如单词或子词），用于训练。训练数据集大小通常以处理的 token 总数来衡量。更大的 token 数量通常反映更多样化和更广泛的训练数据，这可以导致模型性能的提高 – 直到达到某个点 – 然后开始出现收益递减。来源：Epoch AI (5/25)

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

Big Six* USA Public Technology Company CapEx Spend ($B) vs. Global ChatGPT Weekly Active Users (MM) – 2014-2024, per Capital IQ & OpenAI

[1]ChatGPT WAU data as of 11/23 & 12/24 due to data availability.
*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), OpenAI disclosures (3/25)

CapEx Spend @ 六大 * 科技企业 =+63% Y/Y & 加速 …

美国六大 * 公共科技企业 CapEx Spend ($B) 与全球 ChatGPT 周活跃用户 (MM) – 2014-2024, 根据 Capital IQ & OpenAI

[1]ChatGPT WAU 数据截至 11/23 & 12/24 因数据可用性。* 注意：美国六大科技企业包括苹果、英伟达、微软、 Alphabet / Google 、亚马逊 & Meta Platforms / Facebook。仅显示亚马逊的 AWS CapEx & 收入（即不包括亚马逊零售 CapEx ）。AWS CapEx 根据摩根士丹利 – 估算等于 AWS 净增加的房产及设备减去金融租赁和义务。来源：Capital IQ (3/25) 、OpenAI 披露 (3/25)

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

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, Alphaet / 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)

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

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, Alphaet / 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)

BOND CapEx Spend – Big Technology Companies = Inflected With AI's Rise 102

CapEx Spend @ Amazon AWS =

Cloud vs. AI Patterns

CapEx Spend @ Amazon AWS =

云 vs. 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

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)

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

Tech CapEx Spend Partial Instigator =

Material Improvements in GPU Performance

技术资本支出部分驱动因素 =

GPU 性能的重大改进

BOND 105

NVIDIA GPU Performance = +225x Over Eight Years

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

$1B Data Center Comparison
GPT-MoE Inference Workload[1]

	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

[1] *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)*

NVIDIA GPU 性能 =+225x 跨八年间

NVIDIA GPU 系列随时间推移的性能 – 2016-2024, per NVIDIA

$1B 数据中心比较 GPT-MoE
推理工作负载[1]

	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

[1] GPT-MoE 推理工作负载 = 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)

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 TPUs 或其他专用 AI 加速器，因为可用数据较少。TPUs 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

BOND 108

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)

关键技术资本支出受益者 = 英伟达 …25% & 全球数据中心资本支出增长，根据英伟达

全球数据中心资本支出 **($B) 与英伟达数据中心收入的百分比** 数据中心资本支出（全球） **–** 2022-2024，根据英伟达 @ GTC

注意：英伟达数据代表财年 1 月（例如，*2024* = FY25 结束于 1/25），而数据中心资本支出为日历年。由杰森 · 黄在英伟达 GTC 2025（链接）上发布的数据。来源：*Dell'Oro* 研究用于资本支出（3/25）；英伟达用于数据中心收入（3/25）

Technology Company Spend =

R&D Rising Along with CapEx

科技公司支出 =

研发支出随资本支出增长

BOND 110

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

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)

研发支出 @ 美国六大 * 科技上市公司 = 占收入的 13%… 与十年前的 9% 相比

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

* 注意：美国六大科技公司包括苹果、英伟达、微软、字母表 / 谷歌、亚马逊和 Meta 平台 / 脸书。亚马逊的研发费用显示为非 AWS，因为公司财务报表中未细分数据；因此收入按同比方式显示。来源：Capital IQ (3/25)

Tech Big Six (USA) =

Loaded With Cash to Spend on AI & CapEx

Tech Big Six (USA) =

Loaded With Cash to Spend on AI & CapEx

BOND 112

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

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)

六大科技巨头 * 产生大量现金 =+263% 增长在自由现金流十年间达到 3890 亿美元 …

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

*注意：美国六大科技公司包括苹果、英伟达、微软、字母表 / 谷歌、亚马逊和 Meta 平台 / 脸书。自由现金流（FCF）计算为经营活动现金流减去资本支出，以标准化定义，因为只有一些公司减去金融租赁，而亚马逊调整了设备出售收益的 FCF。显示的是亚马逊的 FCF，而不是 AWS，因为公司财务报表中没有细分数据。来源：Capital IQ （3/25）

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

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)

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

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)

Tech CapEx Spend Driver =

Compute Spend to Train & Run AI Models

技术资本支出驱动 =

用于训练和运行人工智能模型的计算支出

BOND 115

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.

整体动态非常明确：单位成本降低正在推动整体支出增加。随着推理成本降低，AI 得到更广泛的应用。而随着 AI 应用增多，整体基础设施和计算需求上升 – 再次推高成本。结果是形成了一个增长飞轮，对云服务提供商、芯片制造商和企业 IT 预算都带来压力。

AI 的经济模式正在快速演变 – 但目前，它们仍然由高资本密集度、大规模基础设施和争夺指数级扩张的用例所驱动。

Data Centers =

Key Beneficiary of AI CapEx Spend

数据中心 =

AI 资本支出 (CapEx) 的主要受益者

BOND 117

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…

要深入了解 AI 基础设施的经济状况，观察数据中心建设的速度和规模是有用的。当前的 AI 驱动需求已将数据中心支出推至历史高位。根据 Dell'Oro 研究，2024 年全球 IT 公司数据中心资本支出达到 4550 亿美元，并正在加速。

超大规模计算公司和 AI 首要公司都在投入数十亿美元来建设计算能力 – 不仅用于存储，还用于实时推理和模型训练工作负载，这些工作负载需要密集、高功率的硬件。随着 AI 从实验性转向必要性，数据中心也是如此。根据英伟达联合创始人兼 CEO 黄仁勋的说法，这些 AI 数据中心 …… 实际上就是 AI 工厂。

这场竞赛的速度比许多人预期的要快。最引人注目的例子可能是 xAI's 位于田纳西州孟菲斯市的巨像设施，该设施从一座破败的工厂到一座完全运营的 AI 数据中心仅用了 122 天。正如第 122 页所述，在 75 万平方英尺 – 大约相当于 418 个美国普通房屋的面积 – 它是在通常建造一栋美国房屋所需时间的一半内建成的。

根据英伟达联合创始人兼 CEO 黄仁勋的说法，他们取得的成就是独一无二的，前所未有 …… 也就是说，简直超乎人类 ……

…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.

… 这类时间表不再是例外。凭借预制模块、简化的许可流程，以及电气、机械和软件系统的垂直整合，新数据中心的建设速度更像是消费科技周期而非房地产发展。

但在这速度之下是一个远非简单的资本模式。资本支出（CapEx）由土地、电力供应、芯片和冷却基础设施 – 尤其是随着 AI 工作负载将热能和电力限制远远超出传统企业计算的需求。相比之下，运营支出（OpEx）主要由能源成本和系统维护主导，特别是对于高密度训练集群，这些集群几乎持续处于高负载状态。

收入由计算销售驱动 – 无论是以 AI API 、企业平台费用或内部生产力提升的形式。但投资回报期通常很长，尤其是对于提前布局需求的垂直整合参与者。对于新进入者，变现可能滞后于建设数个季度甚至数年。

然后还有供应链。电力供应正变得越来越关键。变压器、变电站、涡轮机、 GPU 、电缆 —— 这些都不是一夜之间就能生产出来的商品。在这种情况下，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、计算和软件变现的交汇点。

那些能正确处理这件事的公司可能不仅仅是运行服务器 – 它们将塑造未来十年人工智能经济的地理格局。

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

USA Data Center Annualized Private Construction Value ($B) – 1/14-12/24, per USA Census Bureau

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

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

USA Data Center Annualized Private Construction Value ($B) – 1/14-12/24, per USA Census Bureau

注意：所有数据均进行了季节性调整。数据来自美国人口普查局的“已投入建设的建筑价值调查”（VIP 调查），该调查提供美国每月建筑工作总价值的估计。数据已年度化以避免季节性波动。来源：美国人口普查局，该调查提供美国每月建筑工作总价值的估计。数据已年度化以避免季节性波动。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

数据中心新建与现有容量，美国 =+16x 在新建 vs. +5x 在现有 四年间

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)

注意：仅限美国主要市场（弗吉尼亚北部、亚特兰大、芝加哥、菲尼克斯、达拉斯 - 沃斯堡、希尔斯伯罗、硅谷、纽约都会区）。来源：CBRE，《北美数据中心趋势 H2 2024》（2/25）

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

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

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

数据中心建设时间 (xAI 巨人为代理) =122 天 vs. 234 天用于家庭

122 天 = 一个完全可运营的数据中心 – **2024…**
750,000 平方英尺 = 相当于 418 个美国家庭的面积

750,000 平方英尺

我们被告知建造需要 24 个月。所以我们自己接手了项目，质疑一切，移除了所有不必要的东西，并在四个月内达成了目标。

- xAI 网站

122 Days =
一半建成的房屋 – 2024
（平均建设时间 = *234* 天）

1,792 平方英尺

注意：截至 2025 年 1 月，美国中位数家庭尺寸，根据 FRED 数据。巨人在田纳西州孟菲斯市的前 Electrolux 工厂建造。显示单栋建筑的平均建设时间。衡量从现场工作开始到完成的时间。数据报告于 2024 年，但衡量的是 2023 年开始的家庭建设时间。来源：xAI，美国人口普查局，圣路易斯联邦储备银行，维基媒体共享

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

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

We're running the world's biggest supercomputer, Colossus.
Built in 122 days – outpacing every estimate –
it was the most powerful AI training system yet.
Then we doubled it in 92 days to 200k GPUs.
This is just the beginning…

…We doubled our compute at an unprecedented rate, with a roadmap to 1M GPUs. Progress in AI is driven by compute and no one has come close to building at this magnitude and speed.

- xAI Website, 5/25

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)

数据中心计算（xAI 巨像作为代理） =0 至 20 万张 GPU 在七个月内

xAI 巨像 GPU –4/24-11/24, per xAI

We're running the world's biggest supercomputer, Colossus.
Built in 122 days – outpacing every estimate –
it was the most powerful AI training system yet.
Then we doubled it in 92 days to 200k GPUs.
This is just the beginning…

…We doubled our compute at an unprecedented rate, with a roadmap to 1M GPUs. Progress in AI is driven by compute and no one has come close to building at this magnitude and speed.

- xAI Website, 5/25

注意：我们假设截至 2024 年 11 月 30 日有 20 万张 GPU，根据 xAI 的披露，' 我们在 92 天内将 GPU 数量 [翻倍至 200K 张 GPU。] ' xAI 巨像于 2024 年 8 月 30 日在 4 个数据中心大厅运行了首个任务。我们假设在建设开始日期（距离假设的 2024 年 8 月 30 日开业日期 122 天之前）为零张 GPU。来源：xAI（5/25），孟菲斯商会（12/24）

Data Centers =

Electricity Guzzlers

数据中心 =

耗电大户

BOND 124

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)…*

AI 与能源观察 / 斜体引用此处及接下来的两页内容来自 **‘世界能源展望特别报告 —— 能源与 AI’** （链接）来自 IEA （国际能源署） * – 4/10/25

要了解能源基础设施的发展方向，有助于审视 AI 能力和电力供应之间日益增长的紧张关系。人工智能的规模和复杂性的日益增长需要大量的计算能力，主要来自专注于 AI 的数据中心。这些设施 – 专门为训练和提供服务模型而建 – 开始在电力消耗上与传统重工业相媲美。

没有人工智能就没有能源 —— 特别是电力 (第 3 页)。数据中心 占全球电力消耗的约 1.5%，在 2024 年 (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每年增长约 12%，比总电力消耗的速度快四倍以上 (第 14 页)。随着电力需求的上升，其集中度也在提高：美国占 ……*[45%* 的全球数据中心电力消耗 *]*, 其次是中国 (25%) 和欧洲 (15%)…… 美国数据中心容量近一半集中在五个区域集群 (第 14 页)。

同样如此：除中国以外的新兴和发展中经济体占 全球互联网用户的 50%，但占全球数据中心 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.

…AI 的电力需求正在增加 – 并且其进展越来越多地受到电网和与需求相关的压力的限制，而不是数据或算法。

虽然 AI 目前对能源部门提出了相当大的需求，但它也已经开始解锁主要的能源效率和运营收益 ……AI 已经被能源公司部署用于改造和优化能源和矿产供应、电力发电和输电以及能源消耗（第 16 页）。

当前的 AI 驱动需求非常高。这预计将持续下去，特别是随着资本涌入模型提供商，而这些提供商反过来又投入更多计算。在某个时刻，这些模型构建者将需要盈利才能投入更多。

虽然需求 – 对于计算和能源 – 随着消费者和企业使用的增加，不可避免地将继续上升，但数据中心最终只会服务于那些支付账单的人。

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.

所有数据均显示为全球数据，除非另有说明。斜体文字直接引自报告。

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)

数据中心电力消耗，全球 =+3 倍增长，19 年间，IEA 数据

按数据中心类型和设备划分的数据中心能源消耗，全球 – 2005-2024，IEA 数据

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

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

数据中心电力消耗按地区 = 美国领先，根据 IEA

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

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

数据中心电力消耗按地区 – 2005-2024，根据 IEA

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

Outline

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

Outline

1. 似乎变化正在比以往任何时候都快？是的，就是这样
2. AI 用户 + 使用 + 资本支出增长 = 前所未有的
3. AI 模型计算成本高 / 上升 + 每令牌推理成本下降 = 性能趋同 + 开发者使用量上升
4. AI 使用 + 成本 + 损失增长 = 前所未有的
5. 人工智能变现威胁 = 竞争加剧 + 开源势头 + 中国崛起
6. AI 与物理世界加速 = 快速 + 数据驱动
7. 全球互联网用户从一开始就由人工智能驱动 = 我们从未见过的增长
8. AI & 工作进化 = 真实的 + 迅速

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。随着性能前沿推向更大的参数数量和更复杂的架构，模型训练成本正攀升至数十亿美元。

讽刺的是，这场构建最强大通用模型的竞赛可能正在加速商品化和驱动收益递减，因为各方的输出质量趋同，差异化变得更加难以维持。同时，应用 / 使用这些模型 – 被称为推理 – 的成本正在迅速下降。硬件在改进 – 例如，NVIDIA 的 2024 年 Blackwell GPU 每 token 消耗的能量比其 2014 年 Kepler GPU 前辈少 105,000 倍。再加上模型算法效率的突破，推理成本正在暴跌。

推理代表了一种新的成本曲线，和 – 与训练成本 – 它正在下降而不是上升。随着推理变得更便宜和更高效，LLM 提供者之间的竞争压力增加 – 不仅在于准确性，还在于延迟、正常运行时间和每 token 成本 *。过去需要花费美元的东西现在可能只需要几分钱。而几分钱的东西可能很快就会只需要几分之一美分。

影响仍在展开。对于用户（和开发者）来说，这种转变是一种礼物：以更低的单位成本访问强大的 AI。随着终端用户成本的下降，新产品和服务的创建正在蓬勃发展，用户和使用的采用率也在上升。

然而，对于模型提供者来说，这引发了关于盈利模式和利润的真实问题。培训是昂贵的，服务正在变得便宜，定价能力正在减弱。商业模式正在变化。并且，关于一刀切的 LLM 方法的新问题正在出现，现在有更小、更便宜的模型被训练用于定制用例。

提供者会尝试构建横向平台吗？他们会深入专业应用吗？只有时间会告诉我们。从短期来看，很难忽视通用 LLM 的经济状况看起来像是一种具有风险投资规模烧钱的商品业务。

* 每个 token 的成本 = 在语言模型的操作过程中，处理或生成单个 token （一个词、子词或字符）所产生的费用。这是一个用于评估 AI 模型部署的计算效率和成本效益的关键指标，特别是在自然语言处理等应用中。 ** 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 / 上升 +

每 token 推理成本下降 =

性能趋同 + 开发者使用量上升

BOND 131

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，每周期 AI & 斯坦福

前沿 AI 模型估算训练成本 – 2016-2024，每周期 AI & 斯坦福

目前，[AI 模型训练成本]1 亿美元。今天正在训练的模型可能更接近 10 亿美元。对。我认为如果我们达到 100 亿或 1000 亿美元，我认为这将在 2025 年、2026 年，也许 2027 年 …… 发生

…… 我认为 10 亿美元的模型训练，是的，可能在 2025 年某个时候开始。

- Anthropic 联合创始人兼 CEO Dario Amodei (6/24)

注意：成本为估算值。由于缺乏可靠的成本数据，不包括大多数中国模型。来源：每周期 AI 通过 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）；《在好公司》播客（6/24）

AI Model Compute Costs High / Rising

+

<u>Inference Costs Per Token Falling</u>

=

Performance Converging + Developer Usage Rising

AI 模型推理成本高/上升

+

推理成本每token下降

=

性能趋同 + 开发者使用量上升

BOND 133

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 19th-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 AI 计算的轨迹，回顾 PC 软件早期的一个想法有助于我们。‘软件像气体一样 …… 它会填满它的容器，’ 微软前首席技术官纳森 · 迈尔霍尔德在 1997 年说过。AI 也不例外。随着模型变得更好，使用量也在增加 – 而使用量增加，对计算的需求也随之增加。我们看到了这一点：更多的查询，更多的模型，每个任务更多的 token。对 AI 的需求并没有放缓。它正在利用每一个可用的资源 —— 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，它的进步速度比堆栈中的几乎任何其他层都快，并且以前所未有的速度。正如第 136 页所述，NVIDIA 的 2024 年 Blackwell GPU 生成 token 所需的能量比其 2014 年的 Kepler 前代产品少 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 —— 不仅仅是成本降低，更是架构和材料创新的正在重塑硬件层面的可能性。

这些硬件效率的提升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。然而，到目前为止，它们还不够。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这些资源的整体使用量增加。这正在推动对扩大能源生产能力的新关注 – 以及关于电网管理能力的新问题。

我们再次看到，这是技术中永无止境的 ‘啊 - 哈’ 时刻：成本下降，性能提升，使用量增长，三者同步。这一趋势正在人工智能领域重演。

**British economist William Stanley Jevons first observed this phenomenon in 19th-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.’*

AI Inference 'Currency' = Tokens

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.

Additional context: 1MM tokens = ~750,000 words…roughly
- 3,500 pages of a standard book (12-point font, double-spaced)
- 5,000 ChatGPT responses*

*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)

AI Inference 'Currency' = Tokens

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.

附加上下文：1MM tokens =~750，000 字 …… 大约
- 3，500 页标准书籍（12 点字体，双倍行距）
- 5，000 个 ChatGPT 响应 *

*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)

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

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)

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

Note: Kepler released in 2012. NVIDIA 材料标记上述所示截至 2014 年的开普勒性能阈值。来源：英伟达公司概览 (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

Note: Axis is logarithmic; every axis tick represents a 10x price change

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

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

Note: Axis is logarithmic; every axis tick represents a 10x price change

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

AI Cost Efficiency Gains = Happening Faster vs. Prior Technologies

Relative Cost of Key Technologies by Year Since Launch, per OpenAI, John McCallum, & Richard Hirsh

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

自发布以来关键技术的相对成本，每年，来自 OpenAI 、 John McCallum 和 Richard Hirsh

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…

科技的持续灵光一闪 = 成本下降 + 性能提升 → 应用普及率上升 …

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"美国城市所有城市消费者消费价格指数：信息技术、硬件和服务平均"。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 8.64×10^{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)

… 科技的持续顿悟 -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：1 petaFLOP/s- 天表示一个系统在 1 petaFLOP/ 秒（每秒 10^{15} 次浮点运算）的速率下运行 24 小时所完成的计算工作总量，相当于大约 8.64×10^{19} 次运算。该指标通常用于衡量训练机器学习模型所需的大规模计算量。FRED 数据显示 " 城市消费者综合价格指数：信息技术、硬件和服务（美国城市平均）。" 注意，虽然训练计算量不是衡量模型性能的直接指标，但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 模型推理成本高/上升

+

推理成本每单位下降

=

性能趋同 + 开发者使用量上升

BOND 141

AI Model Performance = Converging Rapidly, per Stanford HAI

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)

AI 模型性能 = 快速收敛，斯坦福 HAI 报告

Performance of Top AI Models on LMSYS Chatbot Arena – 1/24-2/25, per Stanford HAI

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进行投票来比较它们，投票哪个答案更好。结果有助于根据人类判断对不同的语言模型表现进行排名。在任何给定月份中，仅在此次比较中显示得分最高的模型。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 Model Compute Costs High / Rising

+

Inference Costs Per Token Falling

=

*Performance Converging + **Developer Usage Rising***

AI 模型推理成本高/上升

+

推理成本每单位下降

=

性能趋同 + 开发者使用量上升

BOND 143

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…

要理解 AI 开发者活动的激增，了解推理成本的异常下降和强大模型的可及性日益增长是有益的。

在 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巨大改进驱动的。过去只有大型公司才能负担得起的东西，现在个人开发者、独立应用构建者、笔记本电脑上的研究人员以及夫妻店员工都能负担得起。

成本崩溃使实验变得廉价，迭代变得快速，并且为几乎任何人提供了产品化的可行性。

与此同时，性能趋同正在改变模型选择的计算方式。顶级前沿模型与较小、更高效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 – 实际性能的差异可以忽略不计。开发者发现他们不再需要为顶级模型支付溢价以获得可靠的输出。相反，他们可以在本地或通过低成本 API 提供商运行较便宜的模型，并实现功能上相似的结果，尤其是在特定任务数据上进行微调时。

这种转变正在削弱现有模型的定价杠杆并为人工智能开发创造公平的竞争环境 …

…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，基础模型的大量涌现创造了一种新的灵活性。开发者现在可以在数十个模型之间选择 – OpenAI 的 ChatGPT，Meta 的 Llama，Mistral 的 Mixtral，Anthropic 的 Claude，Google 的 Gemini，Microsoft 的 Phi，以及其他 – 每个模型在不同的领域都表现出色。有些针对推理进行了优化，其他则针对速度或代码生成。结果是摆脱了供应商锁定。

开发者没有在单一供应商下整合，该供应商可以控制访问或提高价格，而是将他们的努力分散到多个生态系统中。这种选择的多样性正在赋予新一代建设者选择最适合他们技术或财务需求的模型的能力。

正在出现的是开发者 - 主导的基础设施增长飞轮。随着更多开发者构建 AI 原生应用，他们也创造了工具、包装器和库，使其他人更容易跟随。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加速增长。

每一波开发者活动的浪潮都减少了下一波的摩擦，将从想法到原型再到产品的所需时间压缩。在这个过程中，使用人工智能构建的门槛正在崩溃 – 不仅仅是成本，还包括复杂性。这不再仅仅是一次平台迁移。这是一场创造力的爆炸。

技术历史已经证明 – 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反复强调的那样 开发者！开发者！开发者 … (链接)– 获得最一致开发者用户和用户量增长的平台 – 并且能够扩展并持续改进 – 将获胜。

The AI Developer Next Door

隔壁的 AI 开发者

BOND 146

AI Tool Adoption by Developers = 63% vs. 44% Y/Y

Note: 2023 N=89,184; 2024 N=65,437. Respondents are global. Source: Stack Overflow Developer Surveys (5/23 & 5/24-6/24)

开发者采用 AI 工具 =63% vs. 44% 同比增长

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（2023 年 5 月 & 2024 年 5 月 -6 月）

AI Developer Repositories – GitHub = ~175% Increase Over Sixteen Months

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)

AI 开发者仓库 – GitHub =~175% 十六个月增长率

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，开发者在此共享和管理与人工智能项目相关的代码、模型、数据和文档。这些空间支持协作、重用和人工智能工具及资源的分发。分析显示包括具有 500+ 星级的 GitHub 仓库。基础设施 = 工具用于模型服务、计算管理、向量搜索 & 数据库。模型开发 = 框架用于建模 & 训练、推理优化、数据集工程 & 模型评估。应用开发 = 定制的人工智能驱动应用（多种用例）。来源：Chip Hyuen via 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

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 月度处理令牌数 (T) – 5 月 24 日至 5 月 25 日，根据 Google

去年这个时候，我们在产品和 API 上处理了 9.7 万亿个令牌一个月。现在，我们处理了超过 480 万亿个 —— 那是 50 倍多。

- Google I/O 2025 新闻稿，5 月 25 日

注意：在 Google 产品和 API 上的令牌使用情况。根据 Google 在 5 月 25 日的说法，"去年这个时候，我们在产品和 API 上处理了 97 万亿个令牌。现在，我们处理了超过 480 万亿个 —— 那是 50 倍多 …… 超过 700 万名开发者正在使用 Gemini 进行开发，比去年这个时候多了五倍。"来源：Google，《Google I/O 2025：从研究到现实》（5 月 25 日）

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 – 微软 Azure AI 工坊 =+5x 季度处理令牌量同比

Microsoft Azure AI Foundry Quarterly Tokens Processed (T) – Q1:24-Q1:25, per Microsoft

[微软 Azure AI] 工坊是代理和 AI 应用工厂。

它现在被超过 70,000 家企业和数字原住民 – 从 Atomicwork，到 Epic、富士通和 Gainsight，到 H&R Block 和 LG 电子 – 使用来设计、定制和管理他们的 AI 应用和代理。

本季度我们处理了超过 1 万亿个 token，同比增长 5 倍 – 其中仅上个月就创下了 5000 亿个 token 的记录。

- 微软 FQ3:25 财报电话会议，4/25

注意：来源：微软 FQ3:25 财报电话会议（4/25）

AI Developer Use Cases = Broad & Varied

AI Developer Use Cases – 2024, per IBM

Code Generation	Bug Detection & Fixing	Testing Automation	Project / Workflow Management	Documentation

Refactoring & Optimization	Security Enhancement	DevOps & CI / CD Pipelines	User Experience Design	Architecture Design

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 = 广泛且多样

AI 开发者用例 – 2024，根据 IBM

Code Generation	Bug Detection & Fixing	Testing Automation	Project / Workflow Management	Documentation

Refactoring & Optimization	Security Enhancement	DevOps & CI / CD Pipelines	User Experience Design	Architecture Design

注意：CI / CD 管道是持续集成 / 持续部署管道。来源：IBM，《软件开发中的 AI》（2024 年）；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 ***)*** 长路通向盈利

BOND 152

Outline

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

Outline

1. 似乎变化比以往任何时候都快？是的，就是这样
2. AI 用户 + 使用量 + 资本支出增长 = 无前例
3. AI 模型计算成本高 / 上升 + 每个令牌推理成本下降 = 性能趋同 + 开发者使用量上升
4. AI 使用 + 成本 + 损失增长 = 无前例
5. 人工智能变现威胁 = 竞争加剧 + 开源势头 + 中国崛起
6. AI & 物理世界加速 = 快 + 数据驱动
7. 全球互联网用户由 AI 驱动迅速增长 = 前所未见的增长
8. AI 与工作进化 = 真实 + 迅速

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*

这次不一样，我们将依靠规模来弥补，并且 我们将来会找到如何变现我们的用户，这些通常是商业中最大的危险声明。

话说回来，在科技投资中，偶尔它们可以是金子 – 亚马逊、 Alphabet （谷歌）、Meta （脸书）、特斯拉、腾讯、阿里巴巴、 Palantir…

在人工智能领域，这确实可能这次不一样，领导者将依靠规模来弥补，并且能够在将来变现用户。尽管现在，“ 这次不一样 ” 也意味着竞争是前所未有的 …

我们从未见过如此多的创始人驱动 / 辅助（例如苹果）公司 * 市值超过 1 万亿美元 – 大多数拥有毛利率为 +50% 加上自由现金流 – 在相对透明世界中同时攻击同一个机会，加上全球大国之间的高风险竞争 – 中国和美国。

欧内斯特 · 海明威的短语 逐渐地，然后突然 出自《太阳照常升起》，适用于技术转折点。

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).

资助人工智能增长（和亏损）的资金来自拥有大量自由现金流和雄厚资产负债表的大公司，以及来自世界各地的富有且雄心勃勃的资本提供者。

毫无疑问，这种竞争 / 资本 / 创业的动态组合将迅速推动人工智能的发展，一个谜团是哪些商业模式将是最后的幸存者。

* 公司包括 NVIDIA 、 Microsoft 、 Amazon 、 Alphabet （ Google ）、 Meta （ Facebook ） & Tesla

Technology Disruption Pattern Recognition = Hundreds of Years of Consistent Signals

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.

技术颠覆模式识别 = 数百年持续信号

技术颠覆具有长期重复的节奏：早期的狂热、疯狂的资本形成、激烈的竞争，以及 – 最终 – 明确的赢家和输家。

Alasdair Nairn 的《推动市场的引擎》（此处链接）提炼了两个世纪的这种周期，他的观察对当今的人工智能繁荣具有预见性。他的观察要点如下 ……

当铁路取代运河时，有几年的股价强劲增长。1840 年代的泡沫在过度热情的预期和不断变化的经济状况的重量下破灭 ……

… 任何需要巨额资本支出的技术进步，即使最终成功，在早期几年也总是面临令人失望的回报的真实风险 ……

… 任何需要大量资本支出并在较长时间内才能获得回报的技术，始终将是一项高风险的冒险 – 除非，当然，存在某种形式的竞争保护 ……

… 这些竞争斗争的赢家并不总是拥有最佳技术的那些人，而是那些最能清楚地看到一个行业或市场可能如何发展的人 ……

… 公司和企业投资历史中最清楚的一课是，如果没有某种进入壁垒，先发优势可能会迅速丧失 ……

… 在本研究中反复出现的一个主题是，虽然识别任何新技术中的赢家往往危险且困难，但几乎总是更容易识别谁将成为 '输家 '。

AI-Related Monetization =

Very Robust Ramps

AI 相关变现 =

非常稳健的斜坡

BOND 156

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…

要理解 AI 硬件战略的演变，我们将看看芯片设计的控制权是如何从传统供应商转向依赖它们的平台的。多年来，英伟达一直是 AI 硬件堆栈的中心。其 GPU （图形处理单元）成为训练和推理的默认引擎，因其能够在大规模上处理高度并行计算而备受青睐。其专有技术 – 和无与伦比的规模 – 使其成为行业领导者。

这种依赖 – 加上过大的突然需求 – 造成了限制。尽管英伟达的快速 – 和令人印象深刻的 – 规模扩大，但在加速计算的行业狂热中，对英伟达 GPU 的需求超过了供应。超大规模计算器和云服务提供商正在努力改进其供应链以管理漫长的交付时间。

这种转变正在加速定制硅 – 尤其是 ASIC，即专用集成电路的崛起。与 GPU 不同，GPU 被设计为支持广泛的工作负载，ASIC 被专门设计用于以最大效率处理特定的计算任务。在 AI 中，这意味着针对矩阵乘法、标记生成和推理加速的优化硅。

Google 的 TPU （张量处理单元）和亚马逊的 Trainium 芯片现已成为其 AI 堆栈的核心组件。亚马逊声称其 Trainium2 芯片比标准 GPU 实例具有 30-40% 更好的性价比，实现了更具成本效益的大规模推理。这些不是旁支项目 – 它们是关于性能、经济和架构控制的基石性投资 ……

…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.

… 定制芯片也反映了更广泛的努力来管理人工智能基础设施的经济性。正如亚马逊 CEO 安迪 · 贾西在 2025 年初所说，人工智能不必像现在这样昂贵，将来也不会这样。 定制硅是控制这些费用的一个杠杆。

与此同时，一个新的基础设施专家生态系统正在出现以满足这一需求。

CoreWeave 已成为增长最快的云 GPU 供应商之一，将游戏和加密硬件供应链改造成企业级人工智能客户服务。

甲骨文长期以来被视为传统 IT 供应商，将自己重新定位为 GPU 丰富的云平台，并提供人工智能特定的产品。Astera Labs 是一个不太知名但至关重要的参与者，它构建了高速互连，在 GPU 和内存系统之间传输数据时延迟最小 – 这是一个越来越重要的性能约束。

These firms aren't building foundation models，但他们正在构建基础模型所依赖的东西 .

随着计算需求的增加，它们已成为一个市场中速度、可用性和效率等重要差异化因素的基础设施。

AI Monetization =

Chips

AI Monetization =

Chips

BOND 159

AI Monetization…Chips = NVIDIA Quarterly Revenue +78% to $39B Y/Y…

NVIDIA Quarterly Revenue by Business Line ($B) – 1/19-1/25, per NVIDIA

Note: Gaming includes PC & console gaming. Other includes Enterprise / Pro Vis, Auto, & OEM / Other. NVIDIA's fiscal year ends January 31. The figures in the title compare FQ4:25 to FQ4:24. Source: NVIDIA (1/25) via Morgan Stanley

AI Monetization… 芯片 =NVIDIA 季度收入 +78% 增至 390 亿美元同比 …

NVIDIA 按业务线划分的季度收入（亿美元）– 1/19-1/25，数据来源：NVIDIA

Note: Gaming includes PC & console gaming. Other includes Enterprise / Pro Vis, Auto, & OEM / Other. NVIDIA's fiscal year ends January 31. The figures in the title compare FQ4:25 to FQ4:24. Source: NVIDIA (1/25) via Morgan Stanley

…AI Monetization…Chips = NVIDIA Revenue +28x Over Ten Years…Big Six CapEx + R&D +6x

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)

…AI Monetization… 芯片 =NVIDIA 收入 +28 十年增长十倍 …Big Six CapEx + R&D +6x

Big Six* USA 公共技术公司 R&D + CapEx 支出 ($B) vs. NVIDIA 收入 ($B) – 2014-2024，根据 Capital IQ

*Note: Big Six USA 技术公司包括 Apple 、 Nvidia 、 Microsoft 、 Alphabet / Google 、 Amazon 和 Meta Platforms / Facebook。包括 Amazon AWS + 零售的 CapEx，因为 R&D 费用对于这两项业务部门并非定期分开核算。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% to $8.9B Y/Y, per Morgan Stanley

Google TPU (Tensor Processing Unit) 预计销售额 – 2021-2024, 根据摩根士丹利

TPUs 是专门为 AI 设计的。TPUs 是一种专用集成电路（ASIC），一种为单一特定目的设计的芯片：运行构建和运行 AI 模型所需的独特矩阵和向量基于的数学。

我们的第一颗这样的芯片，TPU v1，于 2015 年内部部署，并迅速在谷歌的不同部门获得了成功 …

…'我们以为可能只生产不到 10,000 颗，'我们的机器学习硬件系统首席工程师 Andy Swing 说。'结果我们生产了超过 100,000 颗。'

- Google 声明 , 7/24

* 数据为摩根士丹利研究的估计值。注意：相对于 GPU，ASIC 是为特定任务定制设计的（例如 AI 模型训练），而 GPU 是通用的。来源：谷歌，摩根士丹利，《GenAI Monetization – 评估 ROI 公式》（2/25）

AI Monetization…Chips =
Amazon AWS Trainium* Sales +216% to $3.6B Y/Y, per Morgan Stanley

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)

AI Monetization…Chips =
Amazon AWS Trainium* Sales +216% to $3.6B Y/Y, per Morgan Stanley

Amazon AWS Trainium Estimated Sales – 2024-2025, per Morgan Stanley

AWS Trainium 芯片系列是 AWS 为 AI 训练和推理专门设计的 AI 芯片，以提供高性能同时降低成本 …

AWS Trainium2 芯片提供比第一代 Trainium 高 4 倍的性能 …*[并提供] 30*-40% 更好的价格性能，优于当前基于 GPU 的 EC2 P5e 和 P5en 实例。

- 亚马逊 AWS Trainium 概述，访问于 5/25

注意：相对于 GPU，ASIC 是为特定任务定制设计的（例如，AI 模型训练），而 GPU 是通用的。数据根据摩根士丹利的研究估计。来源：亚马逊 AWS，摩根士丹利，《生成式 AI 变现 —— 评估 ROI 公式》（2/25）研究。

AI Monetization =

Compute Services

AI Monetization =

Compute Services

BOND 164

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 收入 –2022-2024，根据 CoreWeave

我们在 2025 年的开局在多个方面都表现出色。我们强劲的第一季度财务表现标志着一系列里程碑，包括我们的 IPO 、我们与 OpenAI 的重大战略合作协议以及其他客户成功案例，我们的收购 Weights & Biases 以及许多技术成就 …

… 对我们的平台的强劲需求正在加速，因为 AI 领导者寻求高性能的 AI 云基础设施，以支持最先进的应用。我们正在尽可能快地扩展以捕捉这种需求。未来属于 CoreWeave。

- CoreWeave CEO Michael Intrator，5/25

来源：CoreWeave （截至 5/25 ）

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, 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

来源：Oracle，摩根士丹利估计，《AI 基础设施周期将走向何方》（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 赚钱 … 基础设施连接 =Astera Labs 收入 +242 % 增长至 396MM 美元

Astera Labs – 2022-2024，根据 Astera Labs

Astera Labs 实现了强劲的第四季度业绩，收入环比增长 25%，并以 2024 年 242% 的收入增长年增长结束 …

… 我们预计 2025 年将成为突破性的一年，因为我们进入了一个由我们四个产品系列带来的收入驱动的新增长阶段，以支持多样化的客户和平台。这包括我们旗舰 Scorpio Fabric 产品，用于头节点 PCIe 连接和后端 AI 加速器规模扩展集群。

- Astera Labs CEO Jitendra Mohan，2/25

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 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 也有潜力成为一项我们可以提供给其他公司的可销售服务，就像亚马逊网络服务提供更多网络服务一样，即使它最初是一家书店。所以，我真的认为 Dojo 的潜力非常显著。

- 特斯拉联合创始人兼 CEO 埃隆 · 马斯克，4/23

特斯拉 AI 训练能力（相当于 H100 的 GPU ）

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 Monetization =

Data Layer

AI Monetization =

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 Monetization… 数据标注与评估 =Scale AI 收入 +160% 增长至 8.7 亿美元

Scale AI 收入 – 2023-2024, per Scale AI

数据丰富并非默认状态；它是一种选择。这需要汇聚工程、运营和 AI 领域的顶尖人才。我们的愿景是数据丰富，即我们拥有持续扩展前沿大语言模型更多数量级的生产能力。我们不应在达到 GPT-10 时受限于数据。

- Scale AI 联合创始人兼 CEO Alexandr Wang, 5/24

我们看到 2024 年的收入增长了 160%，并且获得了超过 15 亿美元的新业务。

- Scale AI 企业和产品通讯负责人 Joe Osborne，4/25

注意：2023 年数据为估计值，基于 Scale AI 的企业及产品沟通负责人 Joe Osborne 的说法，他表示："我们预计 2024 年收入将同比增长 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… 数据存储 / 管理 / 处理 =VAST Data 生命周期销售额从 0 到 20 亿美元，仅用六年多时间

VAST Data – 1/19-5/25，根据 VAST Data

一切都在加速。AI 进展的速度不断加快，因为模型构建者基于彼此的发现并推动边界不断扩展。虽然我们自 2022 年初就开始谈论思考机器，但过去 12 个月中推理模型的出现意味着思考机器的时代实际上已经到来 …

… 我们 VAST 认为，实现最大潜力的路径是简化并减少需要解决的基本挑战。如果我们能构建一个简单的方法来涵盖几乎所有为 AI 需要的基础设施层，而不妥协 … 客户将极大地受益。

- VAST Data CEO Renen Hallak，5/25

Source: VAST Data, Silicon Angle

AI-Related Cost Ramps Relative to Revenue =

Can Be Head-Turning

与收入相关的 AI 相关成本上升 =

Can Be Head-Turning

BOND 172

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 Monetization – OpenAI = Revenue vs. Compute Expense, per *The Information*

OpenAI Revenue & Compute Expense ($B) by Year – 2022-2024, per *The Information*

注意：2022 年没有计算成本数据。这些数字是基于公开报告和 The Information 报道的估计。来源：The Information （4 月 25 日及之前）（链接，链接，链接，链接，链接和链接）

AI Monetization – Microsoft / Amazon / Alphabet / Meta = CapEx Up…Free Cash Flow Margins Down

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%**

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)

AI Monetization – 微软 / 亚马逊 / 谷歌 / Meta = 资本支出 上升 … 自由现金流利润下降

资本支出、自由现金流利润、收入增长 – 2023-2024 年，根据 Capital IQ

		资本支出	→ 自由现金流利润	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%**

注意：自由现金流（FCF）按经营活动现金流减去资本支出计算以标准化，因为只有部分公司扣除金融租赁，亚马逊调整 FCF 以反映设备出售收益。亚马逊的统计数据涵盖 AWS 和零售业务；FCF 未按子公司细分。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…

高收入增长 + 高现金消耗 + 高估值 + 高投资水平 =

对消费者是好消息 … 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 (4/25 estimated)	63,920 (Last Raise: 3/25)
Anthropic	2,000 (3/25)	18,000 (Last Raise: 3/25)
xAI	Materially North of 100* (4/25)	12,130 (Last Raise: 11/24)
Perplexity	120 (5/25)	1,410 (Last Raise: 5/25)

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)

选择私有 AI 模型公司 – 5/13/25 =~10 亿美元 + 年度收入与 ~950 亿美元融资额 …

基础模型预估收入与融资额 – 5/13/25，据 Pitchbook 、 The Information 、 Bloomberg 、 The Wall Street Journal 及 CNBC

公司	年度化收入（百万美元）	总 Raised 迄今为止 ($MM)
OpenAI	9,200 (4/25 估算)	63,920 (Last Raise: 3/25)
Anthropic	2,000 (3/25)	18,000 (Last Raise: 3/25)
xAI	Materially North of 100* (4/25)	12,130 (Last Raise: 11/24)
Perplexity	120 (5/25)	1,410 (Last Raise: 5/25)

* 据部分媒体报道，截至 4/25，xAI 收入高达 10 亿美元。注意：OpenAI 的年度收入估算基于 The Information 与 Bloomberg 发布的 2024 年及 2025 年收入预估，假设收入线性增长。数据已四舍五入。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

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 (4/25 estimated)	300,000 (3/25)	33x
Anthropic	2,000 (3/25)	61,500 (3/25)	31x
xAI	Materially North of 100* (4/25)	80,000 (3/25)	N/A
Perplexity	120 (5/25)	9,000 (12/24)	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)*

… 选择私有 AI 模型公司 – 5/13/25 = 高估值收入倍数

Foundation Model Estimated Revenue Multiple – 5/13/25, per Pitchbook, *The Information, Bloomberg, The Wall Street Journal* & *CNBC*

Company	年度化收入（百万美元）	最新估值 ($MM)	Revenue Multiple
OpenAI	9,200 (4/25 估算)	300,000 (3/25)	33x
Anthropic	2,000 (3/25)	61,500 (3/25)	31x
xAI	Materially North of 100* (4/25)	80,000 (3/25)	N/A
Perplexity	120 (5/25)	9,000 (12/24)	75x

* 部分媒体报道称，截至 4/25，xAI 收入高达 10 亿美元。注意：OpenAI 的年度收入估算基于 The Information & Bloomberg 发布的 2024 年及 2025 年收入估算，假设线性收入增长。xAI 估值根据埃隆 · 马斯克的说法。数据已四舍五入。据报道，Perplexity 已于 5/14/25 进行融资谈判，估值达 14 亿美元（投后估值）；但由于在发布时尚未最终确定，我们引用其最后一次最终融资轮的信息。来源：Pitchbook (5/25), The Information (链接), Bloomberg (链接 & 链接) & CNBC (链接 & 链接)

Valuation-to-Revenue Multiple – OpenAI = Looks Expensive…

Estimated Enterprise Value / Next 12 Months Revenue Multiple – 5/25, per Capital IQ & *Bloomberg*

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…

估算的企业价值 / 未来 12 个月的收入倍数 – 5/25，根据 Capital IQ & Bloomberg

注意：OpenAI 的数据为估算值。除 OpenAI 以外的公司未来 12 个月的收入倍数是根据 Capital IQ 的共识估算。OpenAI 的 NTM 收入估算值截至 12/24，由于数据可用性原因。来源：Capital IQ (5/15/25), Bloomberg (链接)

…Revenue-per-User Multiple – OpenAI = In-the-Range

Estimated Annual Revenue Per User ($) – 2024, per Capital IQ, Morgan Stanley, Semrush, *The Information* & Company Disclosures

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

Estimated Annual Revenue Per User ($) – 2024, per Capital IQ, Morgan Stanley, Semrush, *The Information* & Company Disclosures

注意：OpenAI 的数据截至 4/25 为估计值。所有其他上市公司数据截至 12/31/24，使用 2024 财年数据。由于数据可用性（保守地假设为月活跃用户数）；其他数据使用周活跃用户数。我们根据 OpenAI 12/24 的披露，假设平均每周活跃 ChatGPT 用户为 3 亿。我们根据公司估计值，预计 2024 年 ChatGPT 收入为 37 亿美元。Alphabet 的月活跃用户数据基于网站流量测量和全球互联网用户数据估计。Meta 在 2023 年第四季度最后报告了应用家族的 MAPs，我们保守地假设自那时以来没有增长。来源：Capital IQ (12/24), The Information (4/25 及之前) (链接 , 链接 , 链接 , 链接 & 链接), Semrush (11/24), Morgan Stanley, 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
日益竞争激烈且资本密集。
AI 技术创造颠覆的周期在历史上有相似之处。

近期的引人注目的事件包括 1997 年苹果公司几乎破产时其市值仅为 17 亿美元 *，现在为 320 亿美元。

亚马逊的 濒临死亡的时刻发生在第四季度：当时它报告了净亏损 -54.5 亿美元，收入为 97.2 亿美元。

创始人兼时任 CEO 杰夫 · 贝索斯在 2000 年的股东报告中指出对许多资本市场的人来说，这一年是残酷的，当然对亚马逊的股东来说也是如此。截至我上次写信给你的时候，我们的股价已经下跌了 80% 以上。在 2001 年第三季度的亏损谷底，其市值为 220 亿美元，同时它支持了 2300 万个活跃客户账户。现在的市值是 2.2 万亿美元。总共，亚马逊在其 1997 年第二季度推出到其第一个实现净利润的年份（2003 年）结束的二十七个季度中损失了 30 亿美元。对于其最近的二十七个季度（2018 年第三季度至 2025 年第一季度），亚马逊的累计净利润为 1760 亿美元。

谷歌的 IPO 文件（2004 年 4 月）指出，在 2004 年第一季度，在仅完成一轮 A 轮融资后，它将收入的 22%（8600 万美元中的 3.9 亿美元）用于资本支出 —— 在当时，这是一个难以理解的高数字。它以 230 亿美元的市值上市，现在为 2.0 万亿美元 ……

* 市值数据截至 1997 年 7 月 1 日。微软在一个月后于 1997 年 8 月 6 日完成了对苹果的投资。注意：当前市值数据显示截至 2025 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.

*Measured as unlevered free cash flow.
Note: Present market capitalization figures are shown as of 5/14/25.

… 优步烧掉了 -$17B* 在 2016 年至 2022 年之间（在此之前有重大损失）才在 2023 年实现首次自由现金流为正。2022 年，它有 131MM 月活跃平台消费者。优步最后一次股权融资是 G 轮。它的完全稀释的 IPO 市值是 $82B，现在 $ 189B。

特斯拉烧掉了 -$9.2B 在 2009 年至 2018 年之间才在 2019 年实现自由现金流为正。在 2009 年至 2018 年之间的十年里，它交付 ~540K 辆汽车累计亏损了 -$5.6B。它在 2010 年上市时市值为 $1.6B。从 2019-2024 年，它然后赚取了 $40B 交付了 6.7MM 辆汽车。它的市值现在是 $ 1.1T。

要记住的是 – 在大多数情况下，当一切尘埃落定 – 一家公司的估值应该代表其未来自由现金流量的现值。上述公司 – 凭借激进的现金消耗 – 严厉地考验了这一前提，基于产品卓越性 / 持续改进建立了大规模数据驱动的网络效应，发展了技术驱动的竞争优势，并最终证明了质疑者的错误。

只有时间会告诉我们，当前的人工智能 aspirants 将在赚钱方程式的哪一方。

* 按未杠杆自由现金流衡量。注意：当前市场资本化数据截至 5/14/25。

Usage + Cost + Loss Growth =

Unprecedented…

*What About **Future** Monetization + Profits?*

使用 + 成本 + 损失增长 =

前所未有的 …

未来的变现 + 利润呢?

BOND 182

AI Monetization Possibilities =

New Entrants & / Or Tech Incumbents?

AI Monetization Possibilities =

新进入者 & / 或 技术现有企业?

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 模型经济可能的走向，可以关注能力与成本之间日益增长的紧张关系。

训练最强大的大型语言模型（LLM）已成为人类历史上最昂贵 / 资本密集型的工作之一。随着性能前沿推向更大的参数数量和更复杂的架构，模型训练成本正攀升至数十亿美元。

讽刺的是，这场构建最强大通用模型的竞赛可能正在加速商品化和驱动收益递减，因为各方的输出质量趋同，差异化越来越难以维持。与此同时，应用 / 使用这些模型 – 被称为推理 – 的成本正在迅速下降。硬件正在改进 – 例如，NVIDIA 的 2024 年 Blackwell GPU 比其 2014 年的 Kepler GPU 前辈每 token 消耗 105,000 倍更少的能量。再加上模型算法效率的突破，推理成本正在暴跌。

推理代表了一种新的成本曲线，并且 – 与训练成本 – 不同的是，它在下降而不是上升。随着推理变得越来越便宜和高效，LLM 供应商之间的竞争压力 – 不仅在准确性上，还在延迟、正常运行时间和每个 token 的成本 * 上增加。以前需要花费美元的东西现在可能只需要几分钱。而几分钱的东西可能很快就会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

… 影响仍在展开。对于用户（和开发者）来说，这一转变是一份礼物：显著降低获取强大 AI 的单位成本。并且随着终端用户成本下降，新产品和服务的创造正在蓬勃发展，用户和使用的采用率也在上升。

然而，对于模型提供者来说，这引发了关于变现和利润的真实问题。培训成本高昂，服务正在变得便宜，定价能力正在减弱。商业模式正在变化。并且关于一刀切 LLM 方法的新问题正在出现，现在有更小、更便宜的模型正在为定制用例进行训练 *。

此外，传统的商业护城河正在被颠覆。只需看看谷歌。该公司去年 5 月推出了 AI 概览 – 它们位于许多谷歌搜索结果上方。该公司强调截至 4 月 25 日，其 AI 概览 MAU 达到 1.5B…… 值得注意的是，在过去几周里，谷歌开始在部分 AI 概览中添加广告。。

服务提供商将尝试构建横向平台吗？他们会深入专业应用程序吗？一个或两个领导者将推动主导的用户和使用份额及相关变现，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？只有时间会告诉我们。在短期内，很难忽视通用 LLM 的经济性看起来像是有风险投资规模的烧钱型商品业务。

* 例如，OpenEvidence

AI Company Landscape = Varying Degrees of Vertical Integration

Specializations of Ten Leading AI Companies – 4/25, per *The Wall Street Journal*

Developing Models & Chatbots
All 10 of these companies are building generative-AI tools that can create content including text, images and video.

+

Building AI Infrastructure
These seven companies – both tech giants and AI upstarts – are also building the hardware and data centers that provide the power and infrastructure needed to run AI systems.

+

Providing AI Cloud Services
The top cloud providers offer platforms that help businesses leverage AI tech in their own products and workflows.

*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)

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*

开发模型和聊天机器人这 10 家公司都在构建生成式 - 可以创建包括文本、图像和视频内容的 AI 工具。

+

构建 AI 基础设施
这七家公司 – 既有科技巨头也有 AI 初创公司 – 还在构建提供运行 AI 系统所需的硬件和数据中心。

+

提供 AI 云服务
顶级云服务提供商提供平台，帮助企业在其自身产品和工作流程中利用 AI 技术。

* 与甲骨文、软银和 MGX 合作，建设提议中的 Stargate 数据中心网络。来源：华尔街日报，《人工智能革命究竟有多大，四张图表说明》（4/25）

AI Monetization Possibilities =

New Entrants & / Or Tech Incumbents?

AI Monetization Possibilities =

新进入者 & / 或 技术寡头?

AI – New Entrants =

Rapidly Laying Groundwork

AI – 新进入者 =

快速奠定基础

BOND 188

AI Monetization…Foundation Models = Consumer Subscription Models Driving Monetization…

OpenAI ChatGPT, xAI Grok, Google Gemini, Anthropic Claude & Perplexity Consumer Pricing – 5/25, per Companies

OpenAI ChatGPT
$0 (Free) / $20 (Plus) / $200 (Pro)
per Month

xAI Grok
$0 (Free) / $3 (Basic) / $8 (Premium) /
$40 (Premium+) per Month[1]

Google Gemini
$0 (Free) / $19.99 (AI Pro) /
$250 (AI Ultra) per Month

Anthropic Claude
$0 (Free) / $17[2] (Plus) / $100 (Max)
per Month

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)

AI Monetization… 基础模型 = 消费者订阅模式推动变现 …

OpenAI ChatGPT, xAI Grok, Google Gemini, Anthropic Claude & Perplexity 消费者定价 – 5/25, 每公司

OpenAI ChatGPT
$0 (Free) / $20 (Plus) / $200 (Pro)
per Month

xAI Grok
$0 (免费)/ $3 (基础) / $8 (高级) / $40 (高级 +) 每月 [1]

Google Gemini
$0 (免费) / $19.99 (AI P ro) /
$250 (AI Ultra) per Month

Anthropic Claude
$0 (Free) / $17[2] (Plus) / $100 (Max)
per Month

Perplexity
$0 (Free) / $20 (Pro)
per Month

注意：不包括企业计划。1. Grok 定价与 X 高级订阅捆绑。X 高级订阅包括超出 Grok 使用限制改进的额外福利。2. 提供年度折扣。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

OpenAI ChatGPT
From $0.40 (GPT-4.1 nano) to $40 (o3)
per 1MM Output Tokens

xAI Grok
$0.50 (grok-3-mini-beta) to $25 (grok-3-fast)
per 1MM Output Tokens

Google Gemini
$0.15 (1.5 Flash-8B) to $15 (2.5 Pro Preview)
per 1MM Output Tokens[1]

Anthropic Claude
From $1.25 (Claude 3 Haiku) to $75 (Claude 3 Opus)
per 1MM Output Tokens

Perplexity
$1 (Sonar) to $15 (Sonar Pro)
per 1MM Output Tokens

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)

…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

OpenAI ChatGPT
从 0.40 美元（GPT-4.1 nano）到 40 美元
（o3）每 1MM 输出令牌

xAI Grok
$0.50 (grok-3-mini-beta) to $25 (grok-3-fast)
per 1MM Output Tokens

Google Gemini
$0.15 (1.5 Flash-8B) to $15 (2.5 Pro Preview)
per 1MM Output Tokens[1]

Anthropic Claude
From $1.25 (Claude 3 Haiku) to $75 (Claude 3 Opus)
每 1MM 输出令牌

困惑度
$1 (Sonar) 到 $15 (Sonar Pro)
每 1MM 输出令
牌

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)

AI – New Entrants =

Rapid Revenue Growth

AI – 新进入者 =

快速收入增长

BOND 191

AI Monetization – Foundation Models = OpenAI Revenue +1,050% Annually to $3.7B

ChatGPT Paid Subscribers (MM) & Revenue ($B) – 10/22-4/25, per OpenAI & *The Information*

Source: OpenAI disclosures (as of 4/25), The Information (4/25) (link, link, link & link)

AI Monetization – 基础模型 =OpenAI 收入 +1, 每年增长 50% 至 37 亿美元

ChatGPT 付费订阅者 (百万) & 收入 ($B) – 10/22-4/25, 根据 OpenAI & The Information

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 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…

… 我们在数学和计算机科学竞赛问题上优化得相对较少，而是将重点转向更能反映企业实际使用 LLMs 的真实世界任务。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 Monetization – Generative Search = Perplexity Annualized Revenue +7.6x to $120MM in Fourteen Months

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)

AI 赚钱 – 生成式搜索 =Perplexity 年化收入 +7.6x 至 1.2 亿美元在十四个月内

Perplexity: 生成式搜索 – 3/24-5/25, per Perplexity & Bloomberg

Perplexity 最适合被描述为一个答案引擎。你向它提问，它就会给出答案。不同之处在于，所有答案都有来源支撑。这就像学术写作论文 … 人类的特殊性在于我们充满好奇心。我们需要利用 AI 的力量进一步扩展并发现更多知识。

- Perplexity Co-Founder & CEO Aravind Srinivas, 6/24

如果获取信息感觉像与个人研究助理交谈会怎样?

- Perplexity 联合创始人兼 CEO Aravind Srinivas，5/25

注意：3/24 年化收入数据为 Perplexity 联合创始人兼 CEO Aravind Srinivas 3/25 在领英发布的内容中估计的 'Perplexity 已突破 1 亿美元年化收入 … 同比增长 6.3x 且仍高度未商业化。' 来源：Lex Fridman 播客 (6/24), 加州大学伯克利分校 (5/25), 领英 (3/25), Bloomberg, 'AI 初创公司 Perplexity 接近 14 亿 美元估值融资 ' (5/25) (链接)

AI Monetization – Enterprise Search + Agents = Glean Annualized Revenue +10x to $100MM in Twenty-Four Months

Glean – FQ4:23-FQ4:25, per Glean

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)

Note: Glean's fiscal year ends in January. Source: Glean (2/25, 11/24)

AI Monetization – 企业搜索 + 代理 =Glean 年度收入 +10 倍至 $100MM 在二十四个月内

Glean – FQ4:23-FQ4:25，根据 Glean

我们很荣幸帮助世界上一些最大的公司采用 AI 来改造他们的业务。为了真正解锁新的创造力、生产力和运营效率水平，AI 需要借鉴组织知识全貌 —— 并且需要被每个人访问。

你不需要成为提示工程专家来找到答案、生成内容并使用 AI 自动化工作。

- Glean 联合创始人兼 CEO Arvind Jain (9/24)

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

Top 100 AI Companies vs. Top 100 Saas Companies
Median Time to Annualized Revenue Milestone ($MM) – 2018 vs. 2024, per Stripe

Source: Stripe Annual Letter (2/25)

AI Monetization – 2024 vs. 2018 =35% 更快的增长至 $5MM ARR vs. SaaS 对比，根据 Stripe

Top 100 AI Companies vs. Top 100 Saas Companies
Median Time to Annualized Revenue Milestone ($MM) – 2018 vs. 2024, per Stripe

Source: Stripe Annual Letter (2/25)

AI Monetization Possibilities =

New Entrants & / Or **Tech Incumbents**?

AI Monetization Possibilities =

新进入者 & / 或 技术寡头?

AI – Tech Incumbents =

Broad & Steady Product / Feature Rollouts

AI – 技术巨头 =

广泛且稳定的 产品 / 功能 发布

BOND 198

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…

… 科技巨头在设备 & 平台上拥有数十亿全球用户，并持续推出 AI 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 成员	**rs & 400MM+ Office 365 Paid Seats**
Spotify 用户	**678MM**
Amazon	**600MM+ Alexa Devices & 200MM+ Prime Subscribers**
X Users	**500MM+**
Canva Users	**230MM+**

*Meta 包括 Facebook 、 Instagram 、 WhatsApp 、 & Messenger。 **Apple 包括 iPhones 、 iPads 、 Macs 、 & 其他全球 Apple 设备。 *** 截至 2021 年；无更近的公司数据。注意：一些数据是基于过去公司披露 & 网络流量 / 购买历史分析估计的。不同公司可能基于频率 differently 定义 ‘用户’。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 披露 (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)

Tech Incumbent AI Rollouts = Canva – Background Remover & Magic Media (12/19)

Canva Background Remover & Magic Media – 2023-2024, per Canva

我们社区最受欢迎的 Canva 功能之一是一键式图片背景移除，于 2019 年 12 月 …*[推出]* 取得了巨大成功和社区喜爱。- Canva 宣布稿，9/22

Magic Media 让您通过观看文字转化为令人惊叹的、独一无二的图像 – 现在还包括视频和图形 …In less than a year since launching Magic Media's text to image, we've been overwhelmed by our community's enthusiastic response, withalmost 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 宣布稿，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 AI 推出 =Spotify – AI DJ (2/23)

Spotify AI DJ – 2/23-5/25, 根据 Spotify 的说法

AI DJ 和音乐视频 …… 真的是移动平均 …… 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，人们在 Spotify 上的感受，他们说他们喜欢 Spotify 的地方。

- Spotify 联合创始人兼 CEO 丹尼尔 · 艾克，11/24

2018 年，我们内部说过一句话，至今仍然适用：机器学习 – 当时大多数人称之为 AI 的 – 是产品 …… AI 真的是进化的下一步，机器学习允许个性化，AI 也允许在您的数据之上进行实时交互和推理。

- Spotify 联合总裁、首席产品与技术官
Gustav Söderström，4/25

来源：公司公告 (2/23, 5/23, 8/23, 11/24, 4,25, 5/25)

Tech Incumbent AI Rollouts = Microsoft – Copilot (2/23)

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

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)

Tech Incumbent AI Rollouts = Microsoft – Copilot (2/23)

Microsoft: Copilot – 8/23-12/24，根据微软

为了赋予人们发现乐趣、感受创作奇迹并更好地利用世界知识的能力，今天我们通过改进数十亿人每天都在使用的工具 —— 搜索引擎和浏览器，来提升世界从互联网中获益的方式。

今天，我们推出了一款全新的、 AI 驱动的 Bing 搜索引擎和 Edge 浏览器，现已可在 Bing.com 预览，以提供更好的搜索、更完整的答案、新的聊天体验和生成内容的能力。我们认为这些工具是网络的 AI 助手。

- 官方微软博客，2/23

注意：我们假设启动月份没有用户。我们假设截至 12/24 的累计对话为 15B，这是由于微软在 1/24 宣布了 5B 累计对话，并在 12/24 宣布了 2024 年将增加 10B 对话。我们假设 Verge 在 8/23 宣布的 'Bing Chat 上也有超过 10 亿个对话 ' 完全包含截至该日期的 Copilot 对话量。来源：微软公告与收益报告，Verge 援引微软披露（8/23）

Tech Incumbent AI Rollouts = Meta Platforms – Meta AI (9/23)

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

科技巨头 AI 推广 =Meta Platforms – Meta AI (9/23)

Meta Platforms: Meta AI – 11/23-4/25，根据 Meta Platforms

我预计这将是有史以来高度智能和个性化的 AI 助手首次覆盖超过 10 亿人，我预计 Meta AI 将是那个领先的 AI 助手。Meta AI 已经被更多人使用 ……

…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：我们假设 11/23 时用户为零，根据 Meta 12/23 博客文章的说明，"要与我们的人工智能聊天，请开始一条新消息并在 Instagram、Messenger 或 WhatsApp 上选择"创建 AI 聊天"。它们现在已向美国所有人开放。"来源：Meta Platforms 宣布 & 收益报告

Tech Incumbent AI Rollouts = X – Grok (11/23)

X: Grok – 12/24-4/25, per xAI & 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)*

技术巨头 AI 发布 =X – Grok (11/23)

X: Grok – 12/24-4/25, per xAI & Similarweb

xAI 和 Grok 的使命是理解宇宙。我们想要回答最大的问题。

- xAI 创始人兼 CEO 埃隆 · 马斯克，2/25

使用 Grok 的 AI 效果非常好 …… 重要的是要为 AI 编写良好的价值观，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地追求真理 - 的人工智能。

- xAI 创始人兼 CEO 埃隆 · 马斯克，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)*

Tech Incumbent AI Rollouts = Google – Gemini & AI Overviews (12/23)

Alphabet: Gemini & AI Overviews – 3/25-5/25, per Alphabet & *Business Insider*

AI Overviews embedded in Google Search; @ 1.5B MAUs (4/25)

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

Gemini Chatbot Global MAUs (MM)

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 & AI 概述 – 3/25-5/25，根据 Alphabet & Business Insider

AI Overviews embedded in Google Search; @ 1.5B MAUs (4/25)

我们独特的、完整的 AI 堆栈方法继续是我们增长的核心。本季度非常令人兴奋，因为我们推出了 Gemini 2.5，这是我们最智能的 AI 模型，它在性能方面取得了突破，并且被广泛认为是行业最佳模型。

- Alphabet CEO Sundar Pichai, 4/25

Google Gemini 是一个多模态 AI 模型家族，能够理解和生成包括文本、代码、音频、图像和视频在内的各种类型的数据。

Source: Google Gemini

Gemini 聊天机器人全球月活跃用户数（MM）

注意：Gemini 于 12/23 发布 … 应用于 2/24 发布。Gemini 生态系统中的应用数据。用户数量可能与第三方数据提供商 / 面板（如 Similarweb & Sensor Tower）测量的数据不同，因为它们分别仅测量桌面网站和独立移动应用的访问量。来源：Google 公告（4/25 & 5/25）& Business Insider，'谷歌 Gemini 的使用量飙升，但 ChatGPT 和 Meta AI 等竞争对手仍在将其远远甩在后面'（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 AI 推广 = 亚马逊 – Rufus (2/24)

亚马逊：Rufus – 12/22-3/25，根据亚马逊 & 摩根士丹利估计

我们现在有这么多客户只是使用 Rufus 来帮助他们快速了解某个产品。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 …… 个性化服务正变得越来越好 ……

…… 因此，我们预计在整个 2025 年，当你不确定想买什么并希望从 Rufus 获得帮助时，这种情况的数量将继续增加，并且对客户越来越有帮助。

- 亚马逊 CEO 安迪 · 贾西，2/25

亚马逊北美零售预计商品交易总额（十亿美元），过去 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 推出 =TikTok – 交响乐 AI 助手 (6/24)

TikTok: 交响乐助手 – 1/24-4/25，根据 TikTok & Similarweb

创意在 TikTok 上蓬勃发展。当品牌真正拥抱创意勇气和实验时，他们能够直接与社区对话，并邀请他们加入对话。在 TikTok 世界 2024 上，我们推出了交响乐，这是我们由生成式 AI 驱动的广告解决方案套件 ……

…… 凭借交响乐，所有规模的企业、创作者和机构可以将人类想象力与 AI 驱动的效率相结合，帮助在 TikTok 上扩展内容开发、创意和生产力。研究表明，不仅创建 TikTok 首次广告能将购买意愿提升 +37%，品牌好感度提升 +38%，而且 79 % 的 TikTok 用户更倾向于那些展示出清晰理解如何为该平台创建特定内容的品牌。

- TikTok 声明 , 6/24

注意：包括移动端和桌面网站访问。由于政府限制，中国数据可能存在信息限制。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)

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

注意：统计 iPhone 15 Pro 、 iPhone 15 Pro Max 和 iPhone 16 设备的销售量。数据为估算值。来源：公司公告与投资者文件；IDC 通过摩根士丹利（4/25）

AI – Tech Incumbents =

Rapid Revenue + Customer Growth

AI – 科技巨头 =

快速营收 + 客户增长

BOND 209

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)

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

AI Monetization – ‘AI Product’ = Microsoft AI Revenue +175% to $13B Y/Y

Microsoft AI 产品收入 – 2023-2024，根据微软

我们正在我们的技术栈中创新，并帮助客户释放 AI 的全部投资回报率来把握未来的巨大机遇 …

… 目前，我们的 AI 业务已经超过了一个年收入增长率为 130 亿美元，同比增长 175%。

- 微软 CEO Satya Nadella，1/25

预估微软 AI 产品年度营收（亿美元）

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 服务等。详细的分解数据未在财报电话会议上提供。来源：微软新闻稿，“微软云和 AI 力量推动第二季度业绩”(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，据 xAI 与华尔街日报报道

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

* 部分媒体报道称 xAI 收入高达 10 亿美元，截至 4 月 25 日。来源：xAI (2/25)；华尔街日报，《埃隆·马斯克的 xAI 创业公司在新融资轮中估值 500 亿美元》（11/24）（链接）；CNBC，《马斯克表示将在投资者电话会议中为 xAI ' 赋予适当价值 '，消息人士称》（4/25）（链接）

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 Monetization – AI Services = Palantir USA Commercial Customers +65% to 432 Y/Y

Palantir USA Commercial Customers – Q1:23-Q1:25, per Palantir

我们首次实现了美国商业业务的 10 亿美元年度营收，随着 AIP [人工智能平台] 继续推动美国的新客户转化和现有客户扩张。

- Palantir CFO David Glazer , 5/25

随着 AI 模型的进步和改进，我们继续帮助客户最大限度地利用这些模型在生产环境中的能力，通过本体（Ontology）利用企业内部丰富的上下文。我们在精英执行方面的差异化优势使我们能够为客户交付可量化的卓越表现，不断扩大他们相对于 AI 普及率较低者的优势。

- Palantir CRO & 首席法务官 Ryan Taylor, 5/25

Source: Palantir

AI Monetization Possibilities – Enterprise =

Horizontal Platform
& / Or
Specialized Software?

AI Monetization Possibilities – Enterprise =

水平平台和 / 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 AI 商业化将走向何方，问软件本身正在何处整合是有帮助的。

几十年来，商业软件遵循着熟悉的模式：构建一个专业工具，将其销售给狭窄的用户群，并在垂直领域内进行扩展。这是垂直 SaaS 的时代 – Toast 用于餐厅，Guidewire 用于保险，Veeva 用于生命科学 … 每个工具都解决了一个深刻、狭窄的问题。但随着基础模型和生成式 AI 的兴起，其他人正在争夺这些奖项。

进入水平企业平台 – 将 AI 原生生产力、搜索、通信和知识管理整合到一个统一界面中的水平层。将其视为 Slack 、 Notion 和 ChatGPT 的结合，所有这些都集成在一个平台上。

水平企业平台可能 usher in 一种新的商业化形式：不是通过销售隔离的软件许可证，而是通过为整个堆栈中的智能付费。价值从工具转向结果 – 从 CRM 转向自动化的交易摘要，从服务台转向 AI 驱动的解决流程。

这些横向能力仍处于早期阶段，但它们已经被市场领导者和新进入者 alike 所利用。Microsoft 正在将 Copilot 集成到整个技术栈中。Zoom 和 Canva 正在将 GenAI 层叠到面向用户的工作流程中，而 Databricks 正在将 GenAI 注入其数据和开发者技术栈中。与此同时，像 Glean 这样的初创公司正在押注 AI- 优先工作流程 来挑战套件模型 ……

…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.

… 但专业供应商并非停滞不前。事实上，它们正更快地吸收人工智能 —— 嵌入副驾驶、自动化工作流程，并在专有行业数据上微调模型。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。这使它们在部署特定领域智能方面领先一步 —— 这种人工智能不仅总结会议内容，还能标记监管风险、实时优化定价，或起草符合 FDA 要求的文档。在许多情况下，它们的现有地位成为其优势：它们可以将人工智能作为功能而非产品推出，并在不改变购买行为的情况下将其货币化。

人工智能货币化的下一章可能不是一场赢家通吃的战斗，而是一种融合。横向平台将推动广度，将跨职能的知识拼接在一起；专业平台将推动深度，提供能够理解合规、合同和客户意图的人工智能。

问题不在于平台或专业供应商是否获胜 —— 掌握正确的层级，拥有界面，并捕获日志 ic of work itself.

it's who can abs
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 Monetization Possibilities – Enterprise =

水平平台
& / Or
Specialized Software?

Horizontal Enterprise Platform =

SaaS Incumbents
Or
Large Language Model Challengers?

水平企业平台 =

SaaS Incumbents
Or
大型语言模型挑战者?

BOND 217

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

在财报电话会议中提及 "AI" 的次数 – Q1:20-Q1:24, 根据 Uptrends.ai

来源：Uptrends.ai (6/24), 公司公告 & 投资者文件

Enterprise SaaS Incumbent AI Rollouts = 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

企业 SaaS 现有 AI 推广 =Microsoft GitHub Copilot – 6/22

Microsoft GitHub Copilot – 6/17-6/24，根据 GitHub 、 Microsoft & Wells Fargo 的数据

GitHub Copilot 是迄今为止采用最广泛的 AI 驱动开发者工具。自其通用可用性以来仅两年多，已有超过 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

我们很高兴地看到 GitHub Copilot 和 Visual Studio Code 在发布后的第一周内就获得了超过 100 万的注册用户。截至目前，GitHub 现在拥有 1.5 亿开发者，比过去两年增长了 50%。

- 微软 CEO 萨提亚 · 纳德拉，1/25

注意：GitHub 收入披露不规律；3 个数据点来自公司高管的披露。公开的开发者发布日期显示。GitHub 报告年度化收入；这里，我们将其翻译为季度收入。来源：公司公告 & 投资者文件

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)

Enterprise SaaS Incumbent AI Rollouts = Microsoft 365 Copilot – 3/23

微软 365 Copilot – Q2:23-Q4:24，根据微软 & 摩根士丹利

我们看到所有规模的客户采用速度都在加快，因为我们赢得了新的微软 365 Copilot 客户，并且看到大多数现有的 enterprise 客户回来购买更多席位。当你看在可用第一季度期间购买 Copilot 的客户时，他们在过去 18 个月里集体将席位扩展了 10 倍以上。并且总体上，每天使用 Copilot 的人数，再次，每季度翻了一番以上。

员工也前所未有地与 Copilot 进行互动。使用强度每季度增加了 60% 以上，并且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。微软 365 Copilot 于 3/23 发布，但直到 11/23 才对 enterprise 客户正式开放。来源：公司公告 & 投资者文件，摩根士丹利，'4Q24 预览 – 微软能否为 AI 营收问题增加清晰度？(7/24)

Enterprise SaaS Incumbent AI Rollouts = Adobe Firefly – 3/23

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)

Enterprise SaaS Incumbent AI Rollouts = Adobe Firefly – 3/23

Adobe Firefly – 5/23-4/25，根据 Adobe

Adobe FireFly 视频模型于 2 月发布，这是一个商业安全的生成式 AI 视频模型，受到了品牌和创意专业人士的高度好评 ……

… 用户参与度很高，超过 90% 的付费用户正在生成视频 ……

… 我们对这些新产品的前期兴趣感到非常高兴。其他创意专业人士和创作者的亮点包括，我们的产品中 GenAI 的持续强劲采用，Photoshop GenAI 月活跃用户约为 35%，Lightroom GenAI 月活跃用户为 30%。用户已使用 Firefly 生成了超过 20 亿个资产。

- Adobe 数字媒体总裁 David Wadhwani, 3/25

注意：我们假设启动月份没有用户。Adobe Firefly 于 2023 年 3 月作为公共测试版发布。来源：Adobe 宣布（9/23、10/23、3/24、4/24、10/24、12/24、2/25）

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)

Enterprise SaaS Incumbent AI Rollouts = Atlassian Intelligence – 4/23

Atlassian Intelligence – 12/23-12/24，根据 Atlassian 的说法

今天，超过 100 万月活跃用户正在使用我们的 Atlassian intelligence 功能来解锁企业知识、超级充电工作流程并加速他们的团队协作。这些功能显然在创造价值，因为我们看到 AI 交互量同比增长超过 25 倍 ……

…Atlassian Intelligence *[saw a] 25*x improvement in the 过去一年中使用的功能数量。

- Atlassian 联合创始人兼联合首席执行官 Michael Cannon，2/25

注意：12/23 的用户包括测试用户。我们根据 Atlassian 的披露，假设有 20,000 名用户，因为 Atlassian 表示 ' 近 10% 的 Atlassian 的 265，000+ 客户已经通过我们的测试计划利用了 AtlassianIntelligence。' 来源：Atlassian 公告（4/23、12/23、12/24）

Enterprise SaaS Incumbent AI Rollouts = Zoom AI Companion – 9/23

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

Estimated Zoom Accounts with AI Companion Activated (MM)

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)

Enterprise SaaS Incumbent AI Rollouts = Zoom AI Companion – 9/23

Zoom AI Companion – 9/23-12/24, per Zoom

Zoom AI Companion 的月活跃用户增长率加速至环比 68%，这表明 AI 正在为客户提供真正的价值。Zoom AI Companion 已成为我们转型为 AI- 优先公司 …… 的推动力。

…… 作为 AI Companion 2.0 的一部分，我们增加了高级代理功能，包括记忆、推理、编排，以及与 Microsoft 和 Google 服务的无缝集成。

- Zoom 创始人兼 CEO 艾瑞 · 袁 , 2/25

Estimated Zoom Accounts with AI Companion Activated (MM)

注意：AI Companion MAUs 是基于公司披露的估计数据。截至 2024 年 7 月 30 日，Zoom 披露他们有 120 万个已激活 AI Companion 的账户。在 2024 年第三季度，他们披露了活跃账户环比增长 59%；在 2024 年第四季度，他们进一步披露环比增长 68%。我们假设在启动月份没有用户。来源：Zoom 宣布（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：我们假设启动月份没有用户。来源：Canva 公告（10/23、10/24、5/25）

Enterprise SaaS Incumbent AI Rollouts = Salesforce Agentforce – 9/24

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

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)

Enterprise SaaS Incumbent AI Rollouts = Salesforce Agentforce – 9/24

Salesforce Agentforce –12/24-2/25，根据 Salesforce

我们以 9 亿美元的数据云和 AI ARR 结束了这一年。同比增长 120%。我们从未见过产品增长达到这样的水平，尤其是 Agentforce…

… 上线 90 天后，我们已经拥有了 3,000 名付费 Agentforce 客户，他们正在经历前所未有的生产效率、效率和成本节约水平 …

… 数据云是驱动 Agentforce 的燃料，我们的客户正在投资它。数据云已超过 50 万亿，那是带 T 的万亿次，记录数量同比增长翻倍，因为客户增加了对数据平台的消耗和投资。

- Salesforce 联合创始人兼 CEO 马克 · 贝尼奥夫，2/25

注意：Agentforce 于 9/12/24 公布，但于 10/29/24 正式上线。我们假设启动月份没有用户。来源：Salesforce 宣布（10/24 、 12/24 、 2/25 ）

Horizontal Enterprise Platform =

SaaS Incumbents
Or
Large Language Model Challengers?

水平企业平台 =

SaaS Incumbents
Or
Large Language Model Challengers?

BOND 226

OpenAI ChatGPT = Potential Horizontal Enterprise Platform?...

OpenAI = Next-Gen All-in-One Enterprise Platform?

Microsoft Office Suite	OpenAI ChatGPT
Microsoft 365	ChatGPT
9 Applications	1 Application
400MM Paid Users Over 34 Years *1990-2024*	20MM Paid Users Over **2.5 Years** *11/22-4/25*

Source: Microsoft (1/24), Office365 Pros, OpenAI, The Information (4/25) (link)

OpenAI ChatGPT = Potential Horizontal Enterprise Platform?...

OpenAI = 下一代一体化企业平台?

Microsoft Office Suite	OpenAI ChatGPT
Microsoft 365	ChatGPT
9 Applications	1 Application
400MM Paid Users Over 34 Years *1990-2024*	20MM Paid Users Over **2.5 Years** *11/22-4/25*

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 企业版 – 8/23-3/25，根据 OpenAI 、彭博社、路透社及华尔街日报

自 ChatGPT 发布仅九个月后，我们已看到 80% 以上的财富 500 公司采用它。我们听到了商业领袖的反馈，他们希望以简单且安全的方式将其部署在组织中。ChatGPT 企业版的早期用户正在重新定义他们的运营方式，并使用 ChatGPT 来撰写更清晰的沟通内容、加速编码任务、快速探索复杂商业问题的答案、协助创意工作等。

ChatGPT 企业版移除所有使用限制，并最高可快两倍 *[* 与 ChatGPT 免费版相比 *]*… …ChatGPT 企业版还提供无限访问高级数据分析功能，此前被称为 Code Interpreter。

- ChatGPT 企业版发布声明，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（3/2） 5)

AI Monetization Possibilities – Enterprise =

Horizontal Platform
& / Or
Specialized Software?

AI 赚钱的可能性 – 企业 =

水平平台 & / 或 专业软件?

BOND 229

AI Monetization – Enterprise = Specialized Software Opportunities in Fragmented Markets, per Prosus

AI Monetization – 企业 = 专业软件在碎片化市场的机遇，根据 Prosus

USA Industries by Number of Companies & Market Share – 2024, per Prosus

美国产业按公司数量和市场份额 – 2024，根据 Prosus

Source: Prosus, 'The Timeless Appeal of Vertical SaaS' (3/24)

来源：Prosus，《垂直 SaaS 的永恒吸引力》（3/24）

AI-Enabled Specialized Software @
Large Service Industries =

Growing Very Quickly…

Software Engineering
Product Development
Healthcare
Legal
Customer Service
Financial Services

AI-Enabled Specialized Software
@ 大型服务业 =

增长非常迅速 …

软件工程 产品开发 医
疗 法律 客户服务 金
融服务业

AI-Enabled Specialized Software Companies @ Large Service Industries =

Growing Very Quickly…

Software Engineering

AI-Enabled Specialized Software Companies @ 大型服务业 =

增长非常快 …

软件工程

BOND 232

Specialized AI – Software Engineering (Code Editor) = Anysphere Cursor AI ARR @ $1MM to $300MM in Twenty-Five Months

Anysphere Cursor AI – 3/23-4/25, per 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)

专业 AI – 软件工程（代码编辑器） =Anysphere 光标 AI ARR @ 1 千万美元至 3 亿美元，在二十五个月内

Anysphere Cursor AI – 3/23-4/25, per Anysphere

代码正在发生一些美好的事情 …… 我们的目标是用 Cursor 继续引领这一转变，通过构建一个神奇的工具，终有一天会编写世界上所有的软件 ……

… 在 Cursor 中，寻找正确原语的小时数已被即时答案所取代。机械重构被简化为单个 " 制表符 " 。简洁指令被扩展为可工作的源代码。而千行更改在几秒钟内波及到生命。

- Anysphere 新闻稿（ 8/24 ）

… 我们很高兴地报告，Cursor 现在被数百万程序员用作他们的首选编辑器。我们的专有模型现在生成的代码比世界上几乎所有的 LLM 都要多，每天编辑超过十亿个字符。我们的业务规模庞大且快速增长，已超过 1 亿美元的可重复收入。

- Anysphere Team (8/24 & 1/25)

注意：Cursor 于 4/23 上线。我们将 3/23 作为第一个数据点，假设 ARR 为 0 美元。来源：Cursor / Anysphere (8/24, 11/24 & 1/25), Anysphere 联合创始人兼 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)

AI-Enabled Specialized Software Companies @ Large Service Industries =

Growing Very Quickly…

Product Development

AI 驱动的专业软件公司 @ 大型服务业 =

增长非常快 …

产品开发

BOND 234

Specialized AI – Product Development (No-Code Product-Building) = Lovable ARR +13x to $50MM in Five Months

Lovable – 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*

*Per Creandum website. **From Lovable Co-Founder & CEO Anton Osika's LinkedIn posts & podcast appearances. Source: Lovable (5/25), Creandum (2/25)

专业 AI – 产品开发（无代码产品构建） =Lovable ARR +13x 至 $50MM 在五个月内

Lovable – 12/24-5/25

注意：Lovable 是一个 AI 驱动的应用程序开发平台，它使用户能够通过用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。

这里的机遇是巨大的。我们正站在一个范式转变的边缘，其中构建基于软件产品的障碍将消失。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。

- 弗雷德里克 · 卡塞尔，Creandum，***'支持 Lovable：快速行动并创造事物'，2/25****

*Per Creandum website. **From Lovable 公司 - 创始人兼 CEO 安东 · 奥斯卡的领英帖子及播客访谈。来源：Lovable （5/25），Creandum （2/25 ）

AI-Enabled Specialized Software Companies @ Large Service Industries =

Growing Very Quickly…

Healthcare

AI 驱动的专业软件公司 @ 大型服务业 =

增长非常快 …

医疗保健

BOND 236

Specialized AI – Healthcare (Clinical Conversations) = Abridge @ $50MM to $117MM CARR in ~Five Months

Abridge – 10/24-3/25, per 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)

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)

专业 AI – 医疗保健（临床对话） =Abridge @ 5000 万美元至 1.17 亿美元 CARR 在 ~ 五个月内

Abridge – 10/24-3/25，根据 Abridge 和 The Information

Yazdi Bagli，Kaiser IT 和企业管理服务的执行副总裁，表示他相信 [Kaiser Permanente 的] Abridge 合作伙伴关系是医疗保健中最大的生成式 AI 部署之一 …

… 全国推广包括超过 25,000 名医生和临床医生、 40 家医院以及 600 多家医疗办公室以上 …

… 医生的反馈非常热烈：“ 它拯救了我的婚姻。”
以及：“ 你必须从我冰冷、垂死的手中夺走它。”

- 《福布斯》杂志（ 2/25 ）

我们对与 Kaiser 的合作感到无比自豪 – 其中 Kaiser 的大多数医生都在使用 Abridge 来总结患者就诊情况，迄今为止已完成超过 1000 万次。作为我们最早部署之一，这很好地展示了我们如何与众多医院合作伙伴并肩作战，并借助 Abridge 帮助他们发展。

- Abridge 财务总监 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)

AI-Enabled Specialized Software Companies @ Large Service Industries =

Growing Very Quickly…

Legal

AI 赋能的专业软件公司 @ 大型服务业 =

增长非常迅速 …

法律

BOND 238

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)

Source: Harvey (2/25), The Information estimates (1/25) (link, link), & Business Insider (5/25) (link)

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 的报道

2024 年，我们实现了 4 倍的年度经常性收入（ARR）增长，并将客户从 40 家扩展到 235 家，覆盖了 42 个国家，包括美国前 10 大律师事务所中的大多数。

我们也见证了法律服务行业以前所未有的速度转变。律师们正在以前所未有的速度采用技术，数百年老牌的律所正在尝试新的商业模式，企业正在通过 AI 驱动的流程实现显著的成本节约。变化的速度在 2025 年只会加速。

- Harvey 联合创始人兼 CEO Winston Weinberg & 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)

AI-Enabled Specialized Software Companies @ Large Service Industries =

Growing Very Quickly…

Customer Service

AI 驱动的专业软件公司 @ 大型服务业 =

增长非常快 …

客户服务

BOND 240

Specialized AI – Customer Service (AI Support Agents) = Decagon @ ~$1MM to $10MM ARR in One Year

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)

Note: Source: Decagon (12/23, 10/24, 12/24)

专业 AI – 客户服务（AI 支持代理） =Decagon @ ~$1MM 至 $10MM ARR 在一年内

Decagon – 2023-2024, per Decagon

AI 通常被视为破坏就业，但在 Decagon，我们相信相反。我们的 AI 代理正在提升工作，而不是取代它们 …

… 几年后，每家公司都会有 AI 代理运行他们的客户体验。客户支持人员不再处理常规任务；他们现在正在成为 AI 经理 – 配置、培训和监督处理重复性工作的 AI 代理。

- Decagon 联合创始人兼 CEO Jesse Zhang (10/24)

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 驱动的专业软件公司 @ 大型服务业 =

增长非常快 …

金融服务

BOND 242

Specialized AI – Financial Services (Research & Analysis) = AlphaSense @ ~$150MM to ~$420MM in Two Years

AlphaSense – 2022-2024, per 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)

Source: AlphaSense (3/25)

专业 AI – 金融服务（研究与分析） =AlphaSense @ ~$150MM 至 ~$420MM 在两年内

AlphaSense – 2022-2024，根据 AlphaSense

我们正处于一个转折点，AI 驱动的洞察力不再是一种奢侈，而是一种必需 – 每家公司的市场价值都是其做出的决策总和。

超越 $4 亿 ARR 和我们的快速增长是明确的信号，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。

随着我们规模的扩大，我们的重点仍然在于产品和技术创新，确保我们为客户提供高价值解决方案和尖端的 AI 和智能工作流能力。

- AlphaSense 联合创始人兼 CEO 杰克 · 科科（3/25）

Source: AlphaSense (3/25)

Next AI Use Case Frontiers =

Broad & Varied

Next AI Use Case Frontiers =

Broad & Varied

BOND 244

Next AI Use Case Frontiers = Broad & Varied

Next AI Use Case Frontiers – 5/25

Highlights = Pages 246-247

Medical Discovery & Development	Precision Manufacturing	Multi-Purpose Robotics	Autonomous Scientific Research	Supply Chain Optimization
Cybersecurity & Threat Detection	**Personalized Education**	**Autonomous Finance**	**Environmental & Climate Monitoring**	**Energy Grid Management**

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

下一项 AI 应用案例前沿 = 广泛且多样

下一项 AI 应用案例前沿 – 5/25

Highlights = Pages 246-247

Medical Discovery & Development	Precision Manufacturing	Multi-Purpose Robotics	Autonomous Scientific Research	Supply Chain Optimization
Cybersecurity & Threat Detection	**Personalized Education**	**Autonomous Finance**	**Environmental & Climate Monitoring**	**Energy Grid Management**

注意：列表并不全面。来源：药物开发与发现 = Insilico； 精密制造 = Landing AI； 多功能机器人 = Figure AI； 自主科学研究 = IBM 的 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； 能源电网管理 = Uplight； BOND 分析

Next AI Use Case Frontier – Protein Sequencing = Model Size +290% Annually to 98 Billion Parameters Over Four Years

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)

下一个 AI 应用案例前沿 – 蛋白质测序 = 模型大小 +290% 每年增长至 980 亿参数，历时四年

主要蛋白质测序模型的大小（B 参数） – 2020-2024，根据斯坦福大学 RAISE Health

根据斯坦福大学 HAI （4/25）的数据：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特性的能力，加速了结构生物学和分子工程等领域的研究。这些 AI 驱动的技术通过减少对昂贵、耗时的实验方法的依赖，实现了对蛋白质功能和设计的快速探索。

注意：模型列表可能不完整。来源：斯坦福大学 RAISE Health 通过 Nestor Maslej 等人，《AI 指数 2025 年度报告》，AI 指数指导委员会，斯坦福大学 HAI （4/25）

Next AI Use Case Frontier – Protein Sequencing = Synthetically Generated Protein Data Yields 1,000x Expansion via AlphaFold

下一个 AI 应用案例前沿 – 蛋白质测序 =合成生成的蛋白质数据通过 AlphaFold 扩展了 1,000 倍

214,121
Protein Structures
in PDB (2024)

214MM
Predicted Protein Structures
in AFDB (2024)

Note: AFDB predicted protein structure counts may be higher as of year-end 2024. Source: Google DeepMind, RCSB Protein Data Bank (2024)

214,121
Protein Structures
in PDB (2024)

214MM
Predicted Protein Structures
in AFDB (2024)

Note: AFDB predicted 截至 2024 年年底，蛋白质结构数量可能更高。来源：Google DeepMind，RCSB 蛋白质数据库（2024）

Outline

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

Outline

1. 似乎变化比以往任何时候都快？是的，就是这样
2. AI 用户 + 使用 + 资本支出增长 = 前所未有的
3. AI 模型计算成本高 / 上升 + 每令牌推理成本下降 = 性能趋同 + 开发者使用量上升
4. AI 使用 + 成本 + 损失增长 = 前所未有的
5. **AI Monetization Threats** =
   竞争加剧 + 开源势头 + 中国崛起
6. AI 与物理世界加速 = 快速 + 数据驱动
7. 全球互联网用户在 AI 的驱动下从一开始就迅速增长 = 我们从未见过的增长
8. AI & 工作进化 = 真实的 + 快速的

AI Monetization Threats

=

Rising Competition

+

Open-Source Model Momentum
(& China's Rise)

AI Monetization Threats

=

竞争加剧 +

开源模型势头 (& 中国的崛起)

BOND 249

Rising Competition =

AI Model Releases

竞争加剧 =

AI Model Releases

BOND 250

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、视频和传感器读数包含上下文，而文字本身则 ss,
也像 Google 、 Anthropic 和 xAI 等公司
and pee的研究人员 – 们一样 others –
began extending language models to handle additional signals.

多模态 AI 模型就是其结果。它们将文本、图片、声音和视频嵌入到一个共享的表示中，并能在这些格式中生成输出。一个查询可以引用一个段落和一个图表，而模型可以以语音摘要或注释图像的形式进行响应 – 而无需切换系统。每种新的模态都迫使模型在格式之间对齐意义，而不是针对单一格式进行优化。

这一能力逐步发展而来：OpenAI 的 CLIP 在 2021 年将视觉和语言结合起来；Meta 在 2023 年推出 ImageBind，在 2024 年推出 Chameleon ；到 2024-2025 年，GPT-4o 、 Claude 3 和 Chameleon 等前沿系统已成为完全多模态。每种新的模态都迫使模型在格式之间对齐意义，而不是针对单一格式进行优化。

The payoff is practical.
一名现场工程师可以用手机摄像头对准机械并获得平日的故障诊断；一名临床医生可以将 X 光片附加到笔记中并获得结构化报告草稿；一名分析师可以在单个查询中结合图表、转录文本和音频片段。与纯文本模型相比，多模态系统减少了上下文切换，捕捉了更丰富的细节，并使应用 – 质量控制、辅助技术、内容创作 – 成为可能，在这些应用中，视觉或听觉信息与文字同等重要。

Large-Scale AI *Multimodal** Model Competition = +1,150% Rise in Models Released Over Two Years, per Epoch AI

Multimodal Models – Examples

Large-Scale** Multimodal Models – Releases

*A multimodal AI model is one that can process and integrate multiple types of data, e.g., text, images, audio, or video, to understand and generate outputs across different modalities. **Epoch AI defines large-scale as models where their training compute is confirmed to exceed 10^{23} floating-point operations. An AI system can operate in more than one domain and may be double-counted across pages. Source: Epoch AI via Our World in Data (4/25), OpenAI, DeepSeek, Google

Large-Scale AI *Multimodal** Model Competition = +1,150% Rise in Models Released Over Two Years, per Epoch AI

Multimodal Models – Examples

大型 ** 多模态模型 – 发布

* 一个多模态 AI 模型是指能够处理和整合多种类型数据的模型，例如文本、图像、音频或视频，以跨不同模态理解和生成输出。**Epoch AI 将大规模定义为其训练计算量确认超过 10_{23} 次浮点运算的模型。一个 AI 系统可以跨越多个领域运行，并可能在跨页时被重复计算。来源：Epoch AI via 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

Language Models – Examples

Large-Scale* Language Models – Releases

*Epoch AI defines large-scale as models where their training compute is confirmed to exceed 10^{23} 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

Large-Scale AI *Language* Model Competition = +420% Increase in Models Released Over Two Years, per Epoch AI

Language Models – Examples

大型 * 语言模型 – 发布

*Epoch AI 将大规模定义为训练计算量确认超过 10_{23} 次浮点运算的模型。一个 AI 系统可以跨越多个领域运行，并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、DeepSeek、Google

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

*Epoch AI defines large-scale as models where their training compute is confirmed to exceed 10^{23} 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

Large-Scale AI *Vision* Model Competition = +109% Increase in Models Released Y/Y, per Epoch AI

Vision Models* – 示例

Meta Llama 3.2 – 9/24

Qwen2-VL – 12/24

大规模 * 图像模型 – 发布

*Epoch AI 将大规模定义为训练计算量确认超过 10_{23} 次浮点运算的模型。一个 AI 系统可以跨越多个领域运行，并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)、Meta、阿里巴巴

Large-Scale AI *Speech / Audio* Model Competition = +367% Increase in Models Released Y/Y, per Epoch AI

Speech / Audio Models – Examples

OpenAI GPT 4o Speech – 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

大规模人工智能语音 / 音频模型竞赛 =+367% 同比增长，每个 Epoch AI

语音 / 音频模型 – 示例

OpenAI GPT 4o 语音 – 5/24

Microsoft VALL-E – 1/23

Speech / Audio Models – Releases

注意：一个 AI 系统可以操作在多个领域，并且可能跨页面重复计算。包括未经验证的训练计算量的模型。许多展示的模型是多模态的。来源：Epoch AI (5/25), Microsoft (1/23), OpenAI (5/24), 亚马逊, Pinterest

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

*Epoch AI defines large-scale as models where their training compute is confirmed to exceed 10^{23} 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, Amazon, Pinterest, Pinterest

Large-Scale AI *Video* Model Competition = +120% Increase in Models Released Y/Y, per Epoch AI

Video Models – Examples

OpenAI Sora – 12/24

亚马逊新星短片 – 12/24

根据学术研究，人类大脑的 50% 是为视觉处理而构建的。用户能够通过视觉探索他们的兴趣并采取行动 …… 这一点尤其与 Z 世代相关 …… 他们是在由图像和视频构成的全视觉互联网环境中成长起来的。

- 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确认超过 10_{23} 次浮点运算的模型。一个 AI 系统可以跨越多个领域运行，并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

LLM Competition – Website Visits = OpenAI ChatGPT Biggest @ 5.1B Site Visits…

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 = OpenAI ChatGPT Biggest @ 5.1B Site Visits…

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 & xAI Grok 也在增长 @ 每个达到 196-480MM 访问量

DeepSeek, xAI Grok, Perplexity & Anthropic Claude 全球网站访问量 (MM) –5/24-4/25, per Similarweb

Note: Includes desktop & mobile (non-app) we网站访问量。由于政府限制，中国数据可能存在信息限制。来源：Similarweb (5/25)

LLM Competition – Product Releases During Week of May 19, 2025 = It Wasn't Just Google's Annual I/O Conference

Select AI Product Announcements – 5/19/25-5/23/25, per Google, Microsoft, Anthropic & OpenAI

Gemini

- 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

Copilot

- 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

ANTHROP\C

- Claude Opus 4
- Claude Sonnet 4

OpenAI

- Acquisition of io

Note: Announcements include products that were made immediately-available and forthcoming products. List is non-exhaustive. Source: Google Microsoft, Anthropic, OpenAI (5/25)

LLM Competition – 5 月 19 日至 5 月 23 日期间产品发布，2025 = 这并非仅仅是谷歌的年度 I/O 大会

选择 AI 产品发布通知 – 5/19/25-5/23/25，来自谷歌、微软、Anthropic 和 OpenAI

Gemini

- Gemini 实时摄像头和屏幕共享
- Project Mariner 计算机使用
- 更新了 Gemini 2.5 Flash
- Gemini 2.5 Pro
- 2.5 Flash & Pro 预览的本地音频输出
- Gemini 2.5 Pro 的思考预算
- 深度思考
- Project Astra 功能
- Gemini in Chrome
- Deep Research improvements
- Gemini Agent Mode
- Google AI Pro Subscription
- Google AI Ultra Subscription
- Google Beam
- Google Meet 语音翻译
- 个性化智能回复

Jules

- Imagen 4
- Veo 3
- Lyria 2
- Flow TV
- Project Moohan
- Glasses with Android XR
- '试穿' 实验 Agentic checkout
- Gemini 互动测验 Canvas Create 菜单
- LearnLM 集成到 Gemini 2.5 SDK
- Gemini API 中的 Model Context Protocol (MCP) 定义支持 Gemini
- Diffusion SynthID 检测器 对话式导师
- 原型 Google Live API 音视频输入 & 本地音频输出 对话 Gemma 3n AI 工作室
- 增强 Android Studio Journeys
- Android Studio 版本升级代理 Wear
- OS 6 开发者预览 Gemini 代码辅助 新
- Firebase 功能 Google AI Edge 门户
- Google Vids 增强音频概述 Sparkify 实验

Copilot

- Magentic-UI
- Copilot Studio multi-agent orchestration
- GitHub Copilot asynchronous functioning
- Azure AI Foundry 扩展
- NLWeb
- 模型上下文协议 (MCP) 集成
- Entra Agent ID
- SQL Server 2025
- Windows Subsystem for Linux 开源
- GitHub Copilot Chat 扩展
- Aurora AI 驱动的天气预报

ANTHROP\C

- Claude Opus 4
- Claude Sonnet 4

OpenAI

- io 的收购

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竞争威胁

=

竞争加剧

+

开源模型势头强劲（及中国的崛起）

BOND 260

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^{23} floating point operations.*
***To be made open-source as of 6/30/25, per Baidu.*

要了解 AI 模型开发的方向，有助于考察两种不同的方法 – 闭源和开源 – 是如何演变和分化的。

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.

闭源模型遵循集中化、资本密集型的轨迹。这些模型 – 如 OpenAI 的 GPT-4 或 Anthropic 的 Claude – 在专有系统上使用庞大的专有数据集进行训练，需要数月的计算时间和数百万的支出。它们通常提供更强大的性能和更易用的功能，因此受到企业和消费者的青睐，并且 – 越来越多 – 政府。然而，权衡之处在于不透明性：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.

像 Hugging Face 这样的平台使下载 Meta 的 Llama 或 Mistral 的 Mixtral 等模型变得毫不费力，为初创公司、学者和政府提供了无需数十亿美元预算的前沿级 AI。开源 AI 成为现代科技时代的车库实验室：快速、混乱、全球化和激烈协作。并且中国（截至 2025 年第二季度） – 根据发布的大规模 AI 模型 * 的数量 – 正在引领开源竞赛，2025 年发布了三个大规模模型 – DeepSeek-R1 、阿里云 Qwen-32B 和百度 Ernie 4.5**。

分裂会有后果。开源正在推动主权人工智能计划、本地语言模型和社区主导的创新。与此同时，封闭式模型正在主导消费者市场份额和大型企业采用。我们正在观察两种哲学并行展开 —— 自由与控制、速度与安全、开放与优化 —— 它们不仅塑造着人工智能的工作方式，也决定着谁能使用它。

* 大规模人工智能模型 = 训练计算量确认超过 10^{23} 浮点运算。 ** 将于 6/30/25 开源，根据百度。

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)

封闭式与开源模型 – 月活跃用户（MAUs） = 封闭式模型主导消费者，据 YipitData

全球六大领先 LLM 的月活跃用户（MAUs）估算份额 – 4/25，据 YipitData

*xAI 于 2024 年 3 月开源了 Grok-1 基础模型，但更新的版本和完整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，鉴于该面板具有全球代表性（对中国特定数据的限制），相对比较 / 趋势是具有信息量的。仅限桌面用户。数据计算给定平台上的用户数量，除以所有平台上的用户总数。数据是非去重的（即，使用多个平台的用户可能会被计数两次）。数据测量了数百万全球活跃桌面用户的点击流数据。数据由用户的网页请求组成，并从网络服务 / 应用程序收集，例如 VPN 和浏览器扩展。面板具有全球代表性（对中国特定数据的限制）。用户必须连续两个月属于面板才能被纳入。来源：YipitData（5/25）

Closed vs. Open-Source Models – Compute Investment = Closed Models Higher, per Epoch AI

Training Compute Resources for Open vs. Closed LLMs – 2/18-9/24, per Epoch AI

Source: Epoch AI (11/24)

封闭式与开源模型 – 计算投资 = 封闭式模型更高，每个 Epoch 的 AI

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)

封闭式与开源模型 – 性能 = 差距缩小 … 中国崛起，据 Epoch AI…

MATH Level 5 测试中，开源与封闭式 LLMs 的性能对比（按发布年份） – 6/23-4/25，据 Epoch AI

注意：MATH Level 5 pass@1 指的是 AI 模型在 MATH 基准测试上的准确率，该数据集包含高中竞赛级别的数学问题。Level 5 表示基准中最具挑战性的问题。 'pass@1' 衡量模型是否在第一次尝试中正确解决该问题。来源：Epoch AI (5/25)

…Closed vs. Open-Source Models – Performance = Gap Closing…China Rising, per Artificial Analysis

… 封闭式与开放式 - 源模型 – 性能 = 差距缩小 … 中国崛起，据人工智能分析

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)

注意：分数范围为 100 分。测量的每家公司的模型：OpenAI 的 o1；阿里巴巴的 Qwen 2.5 72B；Meta 的 Llama 3.1 405B；Anthropic 的 Claude 3.5 Sonnet。使用的测试包括 HumanEval、MATH-500、MMLU 和 GPQA Diamond。来源：NBC 新闻的人工智能分析，《为何 DeepSeek 与众不同，三张图表说明》（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
=
开发者使用 AI 的爆炸性增长

BOND 266

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.

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.

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.

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.

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.

闭源模型 – 如 GPT-4 、 Claude 或 Gemini – 在消费者和大
ha型企业中占据了主导地位 es,
largely becau他们利用了早期性能优势、易用性和更广泛的认知。 包装在干净、产品化的界面中，并提供
These models came b可靠输出，只需最少的 se tup.

对于企业，它们承诺为非技术人员提供安全性和易用性。
对于消费者，它们带来了品牌知名度、快速入门和完善的 UX。 这种组合使闭源模型保持在
AI 主流的核心。

但性能领先已不再是理所当然。开源模型正在以许多人未曾预料的速度缩小差距 – 而且用户成本只是其中的一小部分 –。像 Llama 3 和 DeepSeek 这样的模型已经展示了具有竞争力的推理、编码和多语言能力，同时它们可以完全下载、微调和在通用基础设施上部署。

这对开发者来说很重要。与企业买家或最终用户不同，开发者不太关心精致度，而更关心原始能力、定制化和成本效率。而且，历史上一直是开发者 – 比任何其他群体 – 引领了 AI 的使用前沿。最近的趋势越来越明显：越来越多的开发者正在转向低成本、高性能的开源模型，使用它们来构建曾经需要封闭 API 的应用程序、代理和管道。

时间将证明这一优势是否能够超越开发者生态系统。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) 和管理服务，这些因素推动了消费者和大型组织中的采用。但随着开源模型成本性能比的持续改善 – 并且如果支持它们的设施变得更加便捷 – 这些优势可能会开始超越开发者社区。

Developer AI Model Activity = +3.4x Increase in Downloads of Meta Llama in Eight Months

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)

开发者 AI 模型活动 =+3.4x 增长 Meta Llama 八个月的下载量

Meta Llama – 8/24-4/25, 来自 Meta Platforms

我预测 2025 年将成为开源成为人们开发的最大类型模型的一年，我认为这很可能是这种情况。这就是我们整体的想法。

- Meta Platforms CEO Mark Zuckerberg, 5/25

Llama 的支持浪潮非常令人振奋。我们十周前宣布 Llama 4 发布后达到十亿下载量。仅仅十周后，这个数字现在达到了 1.2。如果你看看 Hugging Face （下载正在发生的地方），酷的是这些大多是衍生品。我们有成千上万的开发者在贡献。

- Meta Platforms Chief Product Officer Chris Cox, 5/25

Note: 12/24 disclosur统计了 Llama 及其衍生品的下载量。来源：Meta Platforms (8/24, 12/24, 3/25, 4/25), Stratchery 播客 (5/25)

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 增长在 Hugging Face – 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)

注意：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。可能包括开源和闭源模型。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.*

如第 8 页所述，Meta 首席技术官 Andrew Bosworth 将当前 AI 状态称为我们的太空竞赛，我们讨论的对象，尤其是中国，具有很强的能力 ……

在这个背景下，重要的是要记住太空竞赛的赌注是什么：证明哪个政治系统能够更快地创新并在这个过程中赢得世界的信任。在太空竞赛中胜出在增强美国的战略威慑和巩固西方民主价值观的优越性方面发挥了作用。人工智能的“太空竞赛” 也有可能重塑世界秩序。

中国当然知道这些利害关系。早在 2015 年，“中国制造 2025” 这一中国政府的新倡议，旨在推动该国从低成本制造向关键产业的高价值制造转型，似乎还有几十年之遥。快进到今天，中国在这些战略领域的能力已大幅加速 —— 就像机器人技术、电气化和 “信息技术” 一样 —— 这最好地体现在世界一流的人工智能上。

当前中国的人工智能能力已支撑起战场后勤、目标识别、网络作战和自主决策平台等国家级战略领域。2025 年，中国官方媒体强调人工智能融入非作战支持功能（例如军事医院），而科技部则重申其在战略技术领域 “自主创新” 的承诺。

中国人工智能霸权的影响将是深远的。

正如 OpenAI 的 Sam Altman 在 2024 年 7 月的华盛顿邮报评论文章中指出的那样，如果 *[威权政权]* 在人工智能领域取得领先地位，他们将迫使美国公司和其他国家的公司共享用户数据，利用这项技术开发新的监控本国公民的方式，或创建用于攻击其他国家的下一代网络武器。

…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，美国和中国之间更广泛的经济贸易紧张局势仍在升级，这是由于对战略技术输入的控制权竞争所驱动的。目前，中国仍然是全球主要的 " 稀土 " —— 这些对先进电子设备、国防系统和清洁能源基础设施至关重要的材料 —— 供应商 —— 美国正在努力扭转这种不平衡。同时，美国已将半导体制造回流作为优先事项，在《芯片与科学法案》的支持下，并加强了与盟国（包括日本、韩国和荷兰）的合作关系，以减少对中国供应链的依赖。

台湾在这一动态中继续发挥关键作用。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 —— 全球最先进的半导体代工厂 —— 推动了全球半导体生产，因此对两个国家的战略计算都至关重要。

美国醒悟过来花费了很长时间，但在二十年的无所作为之后，两党都在大声呼吁变革。虽然各自采取了不同的方法（拜登政府的出口管制，特朗普政府的经济民族主义和回流政策），但将尖端技术发展视为国家利益核心部分的趋势是一种受欢迎的调整。

正如参议员约翰 · 科尼恩和马克 · 沃纳在 2020 年关于半导体时所指出的，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 的数据，我们知道中国（PRC）的公司 —— 以及其他 —— 正在不断试图提炼领先美国 AI 公司的模型 …… 与美国政府密切合作以最佳保护最强大的模型免受对手和竞争对手为获取美国技术而进行的努力至关重要。

然而，明确的是，自 21 世纪初中国加入世界贸易组织（WTO）时的热情以来，美国对中国的技术的态度发生了变化。人工智能、半导体和关键矿产，以及一般的技术发展，不再被视为仅仅是经济或技术资产 —— 它们代表了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

公共市场资本化领导者谈论过去三十年
=

非凡的美国势头 … 中国崛起

BOND 273

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)

全球公共市场资本领导者 – 2025 年 5 月 2025 =83% (25 of 30) 美国 Based…

按市值排名的全球上市公司 – 5/15/25，根据 Capital IQ 数据

Rank 2025	公司	总部国家	行业	市值 ($B)
1	Microsoft	USA	软件 / 人工智能	$3,368B
2	NVIDIA	USA	半导体 / 人工智能	3,288
3	Apple	USA	硬件 / AI	3,158
4	Amazon	USA	互联网 / 人工智能	2,178
5	Alphabet (Google)	USA	Internet / AI	1,997
6	Saudi Aramco	Saudi Arabia	能源	1,686
7	Meta Platforms (Facebook)	美国	互联网 / 人工智能	1,619
8	Tesla	USA	汽车 / 人工智能	1,104
9	博通	USA	半导 / 人工智能	1,094
10	伯克希尔哈撒韦	USA	金融	1,093
11	TSMC	台湾 n	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	腾讯	中国	软件 / 人工智能	591
17	Mastercard	USA	Finance	529
18	Netflix	USA	互联网 / 人工智能	501
19	埃克森美孚	USA	能源	468
20	好市多批发	USA	消费产品	448
21	Oracle	USA	硬件 / AI	447
22	宝洁	USA	消费品	381
23	Home Depot	USA	消费品	376
24	Johnson & Johnson	USA	消费品	360
25	SAP	Germany	软件 / 人工智能	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 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)

… 全球公共市场资本领导者 – 十二月， 1995 =53% (16 of 30) 美国

按市值排名的全球公共公司 – 12/31/95，彭博社提供

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	Energy	100
5	Coca-Cola	USA	消费产品	94
6	默克	USA	医疗保健	81
7	丰田	日本	汽车	79
8	Roche	瑞士	医疗保健	78
9	Altria	USA	消费品	75
10	Industrial Bank of Japan	Japan	Finance	71
11	MUFG Bank	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	Dai-Ichi Kangyo Bank	日本	金融	61
15	UFJ Bank	日本	金融	59
16	诺华	瑞士	医疗保健	57
17	宝洁	USA	消费产品	57
18	强生	USA	消费产品	55
19	Microsoft	USA	软件	52
20	Walmart	USA	消费品	51
21	IBM	USA	硬件 / 软件	51
22	DirecTV	USA	Media	49
23	Intel	USA	硬件	47
24	BP	美国 Kingdom	能源	46
25	雀巢	瑞士	消费 er Products	45
26	Mobil	USA	Energy	44
27	PepsiCo	USA	消费品	44
28	AIG	USA	金融	44
29	壳牌	联合 大不列颠	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.

在过去三十年的时间里（1995 年至 2025 年），只有六家公司仍然位于前 30 家市值最高的上市公司中 – 微软 / 沃尔玛 / 埃克森美孚 / 宝洁 / 强生 / 可口可乐。

新进入的公司包括英伟达 / 苹果 / 亚马逊 / 谷歌（Alphabet） / 沙特阿美 / Meta Platforms （Facebook） / 特斯拉 / 博通 / 伯克希尔哈撒韦 / 台积电 / 摩根大通 / 维萨 / 伊利利华 / 腾讯 / 万事达 / Netflix / 好市多批发 / 甲骨文 / 家得宝 / SAP / 美国银行 / 工商银行 / 艾伯维 / Palantir。

1995 年，美国拥有 53% （16 家中的 30 家）的市值最高的公司，2025 年则为 83% （30 家中的 25 家）。日本位居其次，1995 年有 9 家，现在为 0。瑞士紧随其后，1995 年有 3 家，现在为 0。英国有 2 家，现在为 0。

2025 年，新的地理进入者包括中国有 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)

全球公共技术市值领导者 –5 月，2025 =70% (21 of 30) 美国 Based …

全球科技公司按市值排名 –5/15/25，根据 Capital IQ

排名 2025	公司	总部国家	行业	市值 ($B)
1	Microsoft	USA	软件 / 人工智能	$3,368B
2	NVIDIA	USA	半导体 / 人工智能	3,288
3	Apple	USA	硬件 / 人工智能	3,158
4	Amazon	USA	互联网 / 人工智能	2,178
5	Alphabet (Google)	USA	互联网 / 人工智能	1,997
6	Meta Platforms (Facebook)	USA	互联网 / 人工智能	1,619
7	Tesla	USA	汽车 / 人工智能	1,104
8	博通	USA	半导 / 人工智能	1,094
9	TSMC	台湾	半导 / AI	856
10	腾讯	中国	软件 / AI	591
11	Netflix	USA	互联网 / 人工智能	501
12	Oracle	USA	硬件 / 人工智能	447
13	SAP	Germany	软件 / 人工智能	343
14	Palantir	USA	软件 / 人工智能	302
15	ASML	荷兰	半导 / AI	300
16	Alibaba	中国	互联网 / AI	281
17	Salesforce	USA	软件 / 人工智能	279
18	T-Mobile	USA	Telco	273
19	Samsung	S. Korea	Hardware / AI	268
20	Cisco	USA	半导体 / 人工智能	256
21	IBM	USA	硬件 / 人工智能	243
22	中国移动	中国	Telco	241
23	Reliance	India	Telco	216
24	ServiceNow	USA	软件 / 人工智能	214
25	Intuitive Surgical	USA	Health Tech	201
26	AT&T	USA	Telco	197
27	Siemens	Germany	硬件 / AI	194
28	Uber	USA	互联网 / 人工智能	189
29	AMD	USA	半导 / 人工智能	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

… 全球公共 技术市值领导者 –12 月，1995 =53% (16 of 30) 美国

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)

全球科技公司按市值排名 –12/31/95，据彭博

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	Telco	34
10	Panasonic	日本	硬件	34
11	SingTel	新加坡	Telco	34
12	Motorola	USA	硬件	34
13	Hitachi	日本	硬件	33
14	Verizon	USA	Telco	29
15	Toshiba	日本	硬件	26
16	Peraton	USA	软件 / Hardware	25
17	Nynex	USA	Telco	24
18	Sony	日本	硬件	22
19	思科	USA	硬件	21
20	Fujitsu	日本	硬件	20
21	PCCW	香港 ong	电信	20
22	NEC	日本	软件	19
23	Oracle	USA	Hardware	18
24	MCI	USA	Telco	18
25	Sharp	Japan	硬件	18
26	TelMex	墨西哥	Telco	17
27	KDDI	日本	电信	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)

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.

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)

在过去的三十年间（1995 年至 2025 年），仅有五家公司仍位列前 30 家最具价值的全球上市科技公司 – 微软 / 甲骨文 / 思科 / 国际商业机器公司 / 美国电话电报公司。

新进入者包括英伟达 / 苹果 / 亚马逊 / 谷歌母公司（Alphabet）/ Meta Platforms （Facebook）/ 特斯拉 / 博通 / 台积电 / 腾讯 / Netflix / SAP / Palantir / 阿斯麦 / 阿里巴巴 / Salesforce / T-Mobile / 三星 / 中国移动 / 电信 / ServiceNow / Intuitive Surgical / 西门子 / Uber / AMD / Intuit。

1995 年，美国拥有 53% （16 家中的 30 家）最具价值的科技公司，2025 年则为 70% （30 家中的 21 家）。1995 年，日本拥有 30% （30 家中的 9 家）顶尖科技公司，2025 年则为 0。英国 / 新加坡 / 香港 / 墨西哥 / 马来西亚曾拥有 1 家，现已为 0。

2025 年，新的地理进入者包括中国（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，尽管台湾在该名单上只有一家公司 – TSMC –，但该公司生产了全球 80%-90% 的最先进半导体和 62%+ 的全球半导体，截至 2024 年第二季度，根据战略与国际研究中心及 Counterpoint Research 的数据。

人们惊叹于一代人之间可以发生多少变化 …… 互联网连接的兴起是大多数新增长的基础。人工智能的兴起在未来三十年将产生相同类型的影响，但可能更快。

来源：战略与国际研究中心，《美国半导体制造回归战略》（2024 年 2 月）；Counterpoint Research，《全球晶圆代工市场份额：季度报告》（2025 年 3 月）

USA vs. China in Technology =

China’s AI Response Time
Significantly Faster vs. Internet 1995

美国与中国在科技领域 =

中国的 AI 响应时间显著快于 1995 年互联网

BOND 280

AI Large Language Model (LLM) Leadership = USA & China Outpacing Rest of World (RoW), per 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^{23} floating-point operations. Source: Epoch AI via Our World In Data (5/25)

AI 大型语言模型 (LLM) 领导力 = 美国 & 中国超越世界其他地区 (RoW)，根据 Epoch AI

按国家累计的大型 AI 系统 * – 2017-2024，根据 Epoch AI

* 香港是中国的一个特别行政区 (SAR)，不是一个独立国家。注意：Epoch AI 将 AI 模型定义为 ' 大型 - 规模，当它们的训练计算量确认超过 10^{23} 次浮点运算时。来源：Epoch AI 通过 Our World In Data (5/25)

China AI = Rapid Relevance…
DeepSeek R1 – 1/20/25…

Reuters

World Business Markets Sustainability Legal Breakingviews Technology Investigations More

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

The Deepseek logo is seen in this illustration taken Jan. 27, 2025. REUTERS/Dado Ruvic/Illustration/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

Source: Reuters, 'DeepSeek narrows China-US AI gap to three months, 01.AI founder Lee Kai-Fu says' (3/25); China Talk Media (11/24)

中国 AI = 快速相关性 …
DeepSeek R1 – 1/20/25…

Reuters

World Business Markets Sustainability Legal Breakingviews Technology Investigations More

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

The Deepseek logo is seen in this illustration taken Jan. 27, 2025. REUTERS/Dado Ruvic/Illustration/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

我们相信，随着经济的发展，中国应该逐渐成为贡献者而不是搭便车者。在过去 30+ 年的 IT 浪潮中，我们基本上没有参与真正的技术创新。我们习惯了摩尔定律从天而降，在家等待 18 个月出现更好的硬件和软件。这就是规模法则被对待的方式 ……

我们看到的是，中国 AI 不能永远处于跟随地位。我们常说中国 AI 与美国之间有一两年差距，但真正差距是原创与模仿之间的差异。如果这不变，中国将永远只是追随者 – 所以一些探索是不可避免的。

- DeepSeek CEO 梁文方，11/24

Source: Reut们，'DeepSeek 缩小中美 AI 差距至三个月，01.AI 创始人李开复 表示 '（3/25）；中国谈传媒 (11/24)

…China AI = Rapid Relevance…
Alibaba Qwen 2.5-Max – 1/29/25…

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 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

Source: Mashable, 'Meet Alibaba's Qwen 2.5, an AI model claiming to beat both DeepSeek and OpenAI's ChatGPT' (1/25); Alibaba (1/25)

…China AI = Rapid Relevance…
Alibaba Qwen 2.5-Max – 1/29/25…

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，同时在其他评估中也表现出有竞争力的结果，包括 MMLU-Pro。

我们的基础模型在大多数基准测试中表现出显著优势，我们对后续训练技术的进步将使 Qwen2.5-Max 的下一代达到新的高度感到乐观。

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。我们致力于通过扩展强化学习的创新应用来提升大型语言模型的思考和推理能力。

- 阿里巴巴 Qwen 2.5 新闻稿，1/25

Source: Mash能够，'遇见阿里巴巴的 Qwen 2.5，一个声称击败 DeepSeek 和 OpenAI 的 ChatGPT 的 AI 模型'（1/25）；阿里巴巴 (1/2 5)

…China AI = Rapid Relevance…
Baidu Ernie 4.5 Turbo – 4/25/25

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)

… 中国人工智能 = 快速相关性 … 百度 Ernie 4.5 Turbo – 4/25/25

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 基础模型家族中最新的成员。想象一个不仅聪明，而且经济实惠和通用的 AI。以下是它引人注目的原因：

- 多模态能力：它在处理文本、图像甚至视频方面表现出色，成为开发人员的瑞士军刀。

- 成本效益：输入每百万个 token 仅 0.8 元人民币，输出每百万个 token 为 3.2 元人民币，比其前身便宜 80%，并且是主要竞争对手成本的一小部分。它仅比 DeepSeek V3 贵 40%，并且仅比 GPT-4.5 贵 0.2%。

- 高性能：基准测试显示它与 GPT-4.1 匹配，并在大多数多模态任务中优于 GPT-4o – 每次运行都能带来高影响力的结果。

- 百度在 X 上的帖子，4/24/25

来源：路透社，《百度在激烈竞争中推出新人工智能模型》（4/24/25）；百度通过推特，《通过更强大和更经济的新模型超级充电人工智能创新》(4/24/25)

China AI =
LLM Performance Catching Up to USA Models, per Stanford HAI…

中国人工智能 =LLM 性能赶超美国模型，斯坦福 HAI 报告称 …

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)

美国顶尖 AI 模型与中国 AI 模型在 LMSYS 聊天机器人竞技场上的表现对比 – 1/24-2/25，根据斯坦福 HAI & 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进行投票来比较它们，投票哪个答案更好。这些结果有助于根据人类判断对不同的语言模型的表现进行排名。在此比较中，仅显示每个月得分最高的模型。来源：LMSYS via Nestor Maslej 等，《AI 指数 2025 年度报告》，AI 指数指导委员会，斯坦福 HAI （4/25）

…China AI =
LLMs Achieving Performance with Lower Training Costs, per Epoch AI…

LLM Training Cost by Year Released – 2022-2024, per Epoch AI & NBC News

Source: Epoch AI via NBC News, 'Why DeepSeek is Different, in Three Charts' (1/25)

…China AI =
LLMs Achieving Performance with Lower Training Costs, per Epoch AI…

LLM Training Cost by Year Released – 2022-2024, per Epoch AI & NBC News

Source: Epoch AI via NBC News, 'Why DeepSeek is Different, in Three Charts' (1/25)

…China AI = LLMs Increasingly Powered by Local Semiconductors…

… 中国 AI = 大语言模型越来越多地由本地半导体驱动 …

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 60

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

Source: Financial Times, ‘Huawei delivers advanced AI chip ‘cluster’ to Chinese clients cut off from Nvidia’ (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 60

华为已经开始向其在与中国断绝关系后增加订单的客户交付先进的人工智能芯片“集群”…

- 金融时报，4/29/25

来源：金融时报，《华为向与中国断绝关系的客户交付先进 AI 芯片“集群”》（4/29/25）

…China AI = Industrial Robot Installed Base Higher vs. Rest of World…

… 中国 AI = 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

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)

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., '人工智能指数 2025 年度报告,' 人工智能指数指导委员会, 斯坦福 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

Images of Industrial & Humanoid Robots, per *The Wall Street Journal*

Source: The Wall Street Journal (2/18, 5/22, 9/22, 5/25)

机器人 – 工业与仿人 = 在新的规模上创造新数据

Images of Industrial & Humanoid Robots, per *The Wall Street Journal*

Source: The Wall Street Journal (2/18, 5/22, 9/22, 5/25)

China Consumer AI Usage =

DeepSeek Rose Quickly

中国消费者 AI 使用 =

DeepSeek Rose Quickly

BOND 291

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.

要了解生成式 AI 市场的演变趋势，考察提供商在不同地区、渠道和用户偏好上的差异有助于理解。在全球范围内，OpenAI 的 ChatGPT 在桌面和移动用户份额上仍然是明显的领导者。但表面之下，市场正在发生转变。像 Anthropic 的 Claude 这样的平台正在 gaining momentum，而 Google 的 Gemini 继续在增长。xAI 的 Grok 根据 Similarweb 的数据，全球网站访问量环比增长了 +294%—— 使其在 2/25-3/25 期间成为增长最快的 AI 助手。

地理因素也日益成为塑造哪些模型获胜的关键。ChatGPT 在大多数国家占据主导地位 – 除俄罗斯和中国外，ChatGPT 无法在俄罗斯和中国运营，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，中国月活跃用户排名前 10 的 AI 应用都是国内开发的 ……DeepSeek 、 Kimi 、 Nami AI 和 ERNIE Bot 每个都在积累数百万用户。在中国以外，情况则不同，ChatGPT 以巨大优势领先。这种分化非常明显：中国正在涌现本土冠军，而全球平台在其他地区占据主导地位。这反映了监管、语言、文化契合度和平台覆盖范围等方面的差异。

要记住中国如何限制平台访问的基础。Facebook 、 Twitter 、 Google 和 YouTube 自 2010 年或更早以来对中国公民不可用。其他受限平台包括 Instagram 、 WhatsApp 、 Wikipedia 、 Telegram 和 Spotify，以及最近 ChatGPT 、 Google Gemini 、 Anthropic Claude 、 Meta AI 和 Microsoft Copilot。

情绪也各不相同。根据斯坦福 HAI 和 Ipsos 的调查，中国公民对 AI 的净收益比美国同行更乐观。2024 年 83% 的中国受访者表示 AI 产品和服务的好处多于坏处，较 2022 年的 78% 有所上升。相比之下，只有 39% 的美国受访者持相同观点，两年间变化不大。

这也反映了一个更深层次的理念分歧，即社会如何适应 AI：不仅是谁在构建它，而且人们如何感知和接受它。在这种环境下，平台选择不仅仅是关于价格或性能。它可能越来越多地受到国家认同的影响。

LLM User Share – Desktop Users = OpenAI ChatGPT Leads…DeepSeek Rose Quickly, per YipitData…

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)

LLM 用户份额 – 桌面用户 =OpenAI ChatGPT 领先 …DeepSeek 迅速增长，据 YipitData…

Estimated Global Monthly Active Desktop User Share – 2/24-4/25, per YipitData

* 仅限聊天机器人。不包括 Gemini 集成在其他地方。注意：这些五个提供商的用户份额显示；其他 LLMs' 的用户份额未显示。仅限桌面用户。数据计算特定平台上用户数量，除以所有平台上用户总数。数据是非重复的（即，使用多个平台的用户可能被计算两次）。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；然而，鉴于该面板具有全球代表性（中国特定数据有限），相对比较 / 趋势是有意义的。数据衡量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 / 应用程序收集，例如 VPN 和浏览器扩展。面板具有全球代表性（中国特定数据有限）。用户必须连续两个月成为面板的一部分才能被包括。数据是非重复的；即，一些用户可能使用多个平台。来源：YipitData （访问时间：5/25 ）

…LLM User Share – Mobile App Users = OpenAI ChatGPT Leads…DeepSeek Rose Quickly, per Sensor Tower…

LLMs – Global Monthly Active Mobile App User Share – 2/24-4/25, per Sensor Tower

*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)

…LLM 用户份额 – 移动应用用户 =OpenAI ChatGPT 领先 …DeepSeek 快速崛起，据 Sensor Tower 报道 …

LLMs – Global Monthly Active Mobile App User Share – 2/24-4/25, per Sensor Tower

* 仅限聊天机器人。不包括其他集成 Gemini 的地方。注意：这五个提供商的用户份额所示；其他 LLM 的用户份额未显示。由于报告差距，中国数据可能不完整。ChatGPT 应用截至 5/25 不可在中国、俄罗斯和部分其他国家使用。数据是非重复的；即，一些用户可能使用多个平台。仅限独立应用数据。来源：Sensor Tower （访问于 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 + 用户 =ChatGPT 支持强劲势头 …

Global Statistics on Apple App Store + Google Play Store – 2/25-4/25, per Sensor Tower

		下载量 (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
	困惑度	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 集成在其他地方。注意：由于报告差距，中国数据可能不完整。截至 5 月 25 日，ChatGPT 应用程序在中国、俄罗斯和部分其他国家不可用。数据是非重复的；即，一些用户可能使用多个平台。仅限独立应用程序数据。来源：Sensor Tower（访问于 5 月 25 日）*

…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 用户分享 – 查询量 =OpenAI ChatGPT 领先，来自 Google

LLMs – 全球每日查询量（MM） –3/28/25，来自 Google

* 仅限聊天机器人。不包括其他集成 Gemini 的地方。注意：DeepSeek 数据不包括中国使用情况。数字已四舍五入。Meta AI 数据引用为 '>200M。' 来源：Google 在 ' 美国诉 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，及市场情报估计，如 Business Insider 报道，' Google 的 Gemini 使用量飙升，但像 ChatGPT 和 Meta AI 这样的竞争对手仍在遥遥领先 '（4/25）（链接）

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 (MM)，中国 vs. 全球 – 3/25，根据 Roland Berger 咨询

注意：总部 = 总部。两个图表的轴采用不同的比例。来源：罗兰贝格通过 AICPB，'2025 年中国生成式 AI 市场五大趋势'（3/25）；中国国家统计局（1/25）；美国人口普查局（4/25）

AI Benefits vs. Drawbacks – China vs. USA Citizens = China Materially More Optimistic Regarding Benefits

'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)

AI Benefits vs. Drawbacks – 中国 vs. 美国 公民 = 中国 在 优势 方面 显著 更乐观

'Products & Services Using AI Have More Benefits than Drawbacks' – 2022-2024, per Stanford HAI & Ipsos

注意：N = 19,504 名 16-74 岁的在线成年人来自 28 个国家。来源：Ipsos，《AI 监测 2024》（6/24），引自 Nestor Maslej 等，《AI 指数 2025 年度报告》，"AI 指数指导委员会，斯坦福 HAI （4/25）

Outline

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

Outline

1. 似乎变化正在比以往任何时候都快？是的，就是这样
2. AI 用户 + 使用 + 资本支出增长 = 前所未有的
3. AI 模型计算成本高 / 上升 + 每令牌推理成本下降 = 性能趋同 + 开发者使用量上升
4. AI 使用 + 成本 + 损失增长 = 前所未有的
5. 人工智能变现威胁 = 竞争加剧 + 开源势头 + 中国崛起
6. AI & 物理世界加速 = 快速 + 数据驱动
7. 全球互联网用户由 AI 驱动从一开始就迅速增长 = 我们从未见过的增长
8. AI & 工作进化 = 真实的 + 快速

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.

在大多数情况下，我们专注于人工智能的势头和桌面 / 移动软件的变现 …… 人工智能在我们的物理世界中的势头和变现，在某些方面甚至更加引人注目。我们正进入一个智能不仅嵌入在数字应用中，还嵌入在车辆、机器和防御系统中的时代。

除了数字代理的崛起，世界正越来越多地经历物理代理的崛起。

像 Waymo 和 Tesla 的完全自动驾驶（ FSD ）测试车队这样的自动驾驶车队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，记录了数百万英里的无人类驾驶，并且软件循环越来越自主。它们下面的技术栈正在变得更智能，数据也更大、更丰富。例如，Applied Intuition 正在构建模拟平台和软件定义的车辆系统，这些系统将自主性从硬件中抽象出来 – 这样制造商就可以像运送零件一样轻松地运送智能。根据 Uber CEO Dara Khosrowshahi 的说法，快进 15 、 20 年，我认为自动驾驶司机将比人类司机更优秀。他们将接受终生的驾驶训练，这是任何人都无法做到的 他们不会分心。。

我们正在看到物理世界 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样子 – 将配备 AI 的自主无人机和反入侵系统运送到每个边缘节点，而不仅仅是指挥中心。在农业领域，像 Carbon Robotics 这样的公司正在将 AI 埋入土壤 – 利用计算机视觉来消除杂草，而无需使用除草剂。

我们相信这些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。智能，曾经局限于屏幕和仪表板，变得具有动态性。

Physical World AI – Vertically-Integrated Electric Vehicles (Tesla) = ~100x Increase in Fully Self-Driven Miles Over Thirty-Three Months

Tesla Vertically-Integrated Electric Vehicles

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

Source: Tesla Disclosures & Q1:25 Investor Deck

物理世界人工智能 – 垂直整合电动汽车（特斯拉） =~100x 增加在 33 个月内的完全自动驾驶里程

特斯拉垂直整合电动汽车

对于完全自动驾驶，我们发布了版本 12，这是与先前版本相比的完整架构重写。这是端到端的人工智能 ……

…… 而且它真的 …… 相当深刻的差异 ……

… 所以，这是第一次使用 AI，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，作为附注，我认为特斯拉可能是世界上最有效率的人工智能推理公司。出于必要性。

- 特斯拉 CEO 埃隆 · 马斯克，1/24

来源：特斯拉披露 & Q1:25 投资者演示文稿

Physical World AI – Fully-Autonomous Vehicles (Waymo) = 0% to 27% Share of San Francisco Rideshares Over Twenty Months, per YipitData

物理世界人工智能 – 完全自动驾驶汽车（Waymo） =0% 至 27% 的旧金山拼车市场份额，持续二十个月，数据来源：YipitData

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

Waymo 完全自动驾驶汽车

[我们正在创建] 一个端到端、非常非常稳健且规模庞大的端到端系统，其基础是多模态的，以便感知规划与预测… 能够比今天更加稳健。

- Waymo 首席运营官 Tekedra Mawakana，1/25

我们在旧金山所做的是向自己和世界证明 —— —— 不仅自动驾驶是可行的，而且它能在市场上规模化运作，并成为一项可行的商业产品。

- Waymo Co-CEO Dmitri Dolgov, 3/25

预估市场份额（总预订额） – 2023 年 8 月 23 日至 2025 年 4 月 25 日，旧金山运营区，根据 YipitData 数据

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)

注意：数据来源于由 >1mm 个每月进行交易的美国电子邮件账户组成的美国用户电子邮件收据面板，涵盖所有可用域。仅统计付费行程。由于样本量有限，数字为估算值。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

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).*

物理世界 AI – 车辆智能（应用直觉） = 服务汽车、卡车、建筑与国防

应用直觉车辆智能

在过去的几年里，我们看到了人工智能的巨大进步，这将对应用直觉服务的行业产生革命性的影响。我们作为生态系统的领导者，是将硅谷最好的资源带给我们的全球客户群的角色。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 联合创始人兼 CEO Qasar Younis, 3/24

我们看到了我们的人工智能工具、自主软件和车辆操作系统的加速采用，因为传统 OEM 厂商正在看到强劲的投资回报率。国防部门也在寻找车辆智能解决方案。我们已经为国防部门提供越野自主技术堆栈多年，并且在过去一年中显著扩展了我们的国防技术产品组合。

- Applied Intuition 联合创始人兼 CTO Peter Ludwig, 5/25

注意：OEM = 原始设备制造商。来源：应用直觉

应用直觉服务全球顶级汽车 OEM –2016-2024，根据应用直觉 *

* Applied Intuition 为不同垂直领域的广大客户提供服务，例如保时捷 / 丰田（汽车）、特拉顿 / 日野（卡车）、卡特彼勒（建筑）和几个美国军事部门（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

Anduril Estimated Revenue ($MM) – F2020-F2024, per News Reports

Source: Anduril, Forbes, TechCrunch, CNBC

物理世界 AI – 美国国防（安杜瑞尔） =+2x Y/Y 过去两年的收入增长

安杜瑞尔 AI 驱动型美国国防系统

在安杜瑞尔，我们坚信当今最紧迫的国家安全挑战不能没有 AI 驱动型系统和大规模自主性。这些系统将有助于保护我们的服务人员并赋予他们在现代化战争的速度下做出更好的决策 ……

… 当正确开发和部署时，［AI 和自主系统］可以使战争比以往任何时候都更加适度、更加精确、更加不任意。

- Anduril 联合创始人兼 CEO 布莱恩 · 施密夫，12/23

安杜瑞尔预计收入（百万美元） – 2020-2024 年，根据新闻报道

Source: Anduril, Forbes, TechCrunch, CNBC

Physical World AI = AI-Driven Mining Exploration (KoBold Metals) = Reversing Trend in Exploration Inefficiency

KoBold Metals AI-Driven Mining Exploration

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

Source: KoBold Metals, Wired (12/22)

物理世界 AI = AI 驱动的矿产勘探（KoBold Metals） = 扭转勘探效率低下的趋势

KoBold Metals AI 驱动的矿产勘探

我们希望在全球范围内扩大和多样化这些金属的供应，但我们采取了与传统矿产公司完全不同的方法 [。我们团队中有三分之二是软件工程师或数据科学家。

- KoBold Metals 联合创始人兼 CEO Kurt House, 12/22

KoBold 的 机器勘探者技术结合了以前从未使用过的数据集以及传统的地球化学、地球物理和地质数据，在统计关联模型中识别潜力。 KoBold 的 技术通过高效筛选大区域加速勘探，并通过识别最有希望的地点使我们的搜索更有效。

- KoBold Metals Website

Source: KoBold Metals, Wired (12/22)

BOND AI & Physical World Ramps = Fast + Furious 305

Physical World AI – Agricultural Modernization (Carbon Robotics) = 230K+ Acres Weeded / 100K+ Gallons of Glyphosate Prevented

Carbon Robotics AI-Driven Agricultural Modernization

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

Source: Carbon Robotics, Organic Produce Network (12/22), GeekWire (3/25)

物理世界 AI – 农业现代化（碳机器人） =230K+ 英亩除草 / 100K+ 加仑草甘膦预防

碳机器人 AI 驱动农业现代化

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，通过电击分生组织高效识别、定位和清除杂草。该设备每小时可覆盖高达 2 英亩，并射出高达 200,000 棵杂草。

- 碳机器人创始人兼 CEO Paul Mikesell，12/22

我们从农民那里了解到，他们最大的挑战仍然围绕着劳动力和劳动力可用性。如果他们能够，他们会 24/7 全天候运行一切。他们会在整个农业季节的每一分钟运行一切，以尽可能多地完成工作。

- Carbon Robotics 创始人兼 CEO Paul Mikesell，3/25

Carbon Robotics 累计除草面积 (K) – 1/23-5/25，数据来源：Carbon Robotics

来源：碳机器人，有机农产品网络（12/22），GeekWire （3/25）

Physical World AI – Intelligent Grazing (Halter) = +150% Net-New Livestock Collars Contracted Y/Y

Halter AI-Driven Intelligent Grazing

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)

Halter Net New Collars Contracted (K) – 2023-2025*, per Halter

**2025 figures annualized as of Q1:25. Source: Halter (5/25)*

物理世界 AI – 智能放牧（Halter） =+150% 净新增牲畜项圈同比合同

Halter AI-Driven Intelligent Grazing

我们亲眼见证了牧场主对土地和动物的关爱与奉献。我们也看到了农业，这一古老而至关重要的产业，尚未完全享受到现代技术的全部益处。这为牧场主提供了巨大的机遇，以在他们的运营中实现更高的生产力和可持续性。我们相信放牧管理是关键。有效的轮牧制度能够更有效地利用自然资源并提高生产力，同时还能改善土壤健康和根系结构以固碳更多。我们相信更高的生产力不必以可持续性为代价。我们可以为牧场主和地球做贡献。

我们相信放牧管理是关键。有效的轮牧制度能够更有效地利用自然资源并提高生产力，同时还能改善土壤健康和根系结构以固碳更多。我们相信更高的生产力不必以可持续性为代价。我们可以为牧场主和地球做贡献。

- Halter (as of 5/25)

Halter 新增项圈合同 (K) –2023-2025*, 根据 Halter

**2025 年数据截至 2025 年第一季度：来源：Halter（5/25）*

Outline

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

Outline

1. 似乎变化正在比以往任何时候都快？是的，就是这样
2. AI 用户 + 使用 + 资本支出增长 = 前所未有的
3. AI 模型计算成本高 / 上升 + 每令牌推理成本下降 = 性能趋同 + 开发者使用量上升
4. AI 使用 + 成本 + 损失增长 = 前所未有的
5. 人工智能变现威胁 = 竞争加剧 + 开源势头 + 中国崛起
6. AI 与物理世界加速 = 快速 + 数据驱动
7. 全球互联网用户由 AI 驱动从起步就加速 = 我们从未见过的增长
8. AI & 工作进化 = 真实的 + 快速的

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.

得益于低成本卫星驱动的互联网连接 / 接入的增长，2.6B （或占世界人口的 32% ）尚未上网的人群接入互联网的潜力正在增加。这些新用户将从 AI 功能开始。哇！

当这些新用户上网时，他们很可能不会遇到浏览器和搜索栏。他们将从 AI 开始 —— 并用他们的母语。想象一个不涉及在搜索引擎中输入查询的 ‘ 初次体验 ’ 互联网，而是与一个能回应的机器交谈。想象完全跳过传统应用层，由一个代理驱动的界面从一处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。一个以代理为中心的互联网体验可能会颠覆现有的技术等级，去中介化主导平台并重新分配价值。

在这个模型中，赢家不会是那些拥有应用的人，而是那些拥有界面的人。

Global Internet Users = Epic Growth Over Past Thirty-Three Years, 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

注意：由于部分国家的数据存在缺失，2021 年的数据进行了插值处理。地区划分依据联合国定义。在某些年份，部分国家的数据可能无法获取，这可能导致趋势线出现波动或与全球用户数据存在轻微差异。来源：联合国 / 国际电信联盟 (3/25)

Global Internet Penetration = 68% vs. 16% Nineteen Years Ago, per ITU

Source: United Nations / International Telecommunications Union (3/25)

全球互联网普及率 =68% vs. 16%，19 年前，根据国际电信联盟

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

全球互联网普及率按地区划分 @ +70% = 除南亚外所有地区 + 撒哈拉以南非洲，根据国际电信联盟统计

Note: Data unavailable for South Asia region for 2023. 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. Source: United Nations / International Telecommunications Union (3/25)

注意：2023 年南亚地区数据不可用。由于部分国家数据缺失，2021 年数据进行了插值。地区划分根据联合国定义。在某些年份，部分国家的数据可能缺失，这可能导致趋势线出现波动。来源：联合国 / 国际电信联盟（3/25）

Global Internet Penetration by Population Density = 83% of Urban Dwellers Online vs. 48% Rural

Source: United Nations / International Telecommunications Union (3/25)

全球互联网普及率按人口密度划分 =83% 的城市居民在线，而农村为 48%

来源：联合国 / 国际电信联盟 (3/25)

Global Internet Users @ 5.5B = +6% Y/Y & Accelerating, per ITU

Source: United Nations / International Telecommunications Union (3/25)

Global Internet Users @ 5.5B = +6% Y/Y & Accelerating, per ITU

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 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 移动应用 @ 530MM MAUs 在二十三个月 = 全球增长我们前所未见

ChatGPT App Monthly Active Users (MAUs) (MM) – 5/23-4/25, per Sensor Tower

注意：地区依据联合国定义。截至 5 月 25 日，ChatGPT 应用在中国、俄罗斯和部分其他国家不可用。仅包括 Android 、 iPhone & iPad 用户。数据可能低估了真实的 ChatGPT 用户基数（例如，桌面或移动网页用户）。仅针对独立应用的数据。来源：Sensor Tower (5/25)

ChatGPT Mobile App – Top User Countries = India @ 14%...USA @ 9%...Indonesia @ 6%, per Sensor Tower

ChatGPT Mobile App Monthly Active Users (MM), Top 10 Countries – 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 (5/6/25)

ChatGPT 移动应用 – Top 用户国家 = 印度 @ 14%... 美国 @ 9%... 印度尼西亚 @ 6%，来自 Sensor Tower

ChatGPT 移动应用月活跃用户 (MM)，Top 10 国家 – 5/23-4/25，来自 Sensor Tower

注意：地区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不可用。仅包括 Android 、iPhone 和 iPad 用户。数据可能低估了真实的 ChatGPT 用户基数（例如，桌面或移动网页用户）。仅涉及独立应用程序数据。来源：Sensor Tower （5/6/25）

DeepSeek Mobile App @ 54MM MAUs in Four Months = Growth Concentrated in China (34% Users) & Russia (9%)

DeepSeek Mobile App Monthly Active Users (MAUs) (MM) – 1/25-4/25, per Sensor Tower

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)

DeepSeek 移动应用 @ 四个月内 54MM MAUs = 增长集中在 中国 (34% 用户) & 俄罗斯 (9%)

DeepSeek 移动应用月活跃用户 (MAUs) (MM) – 1/25-4/25，依据 Sensor Tower

注意：地区依据联合国定义。仅包括 Android 、 iPhone & iPad 用户。数据可能低估了真实的 DeepSeek 用户基数（例如，桌面或移动网页用户）。仅限独立应用数据。由于信息限制，中国、俄罗斯和部分其他国家的数据可能不完整。来源：Sensor Tower (5/6/25)

New Internet User Growth =

Enabled by AI + Satellites

新互联网用户增长 =

由人工智能 + 卫星支持

BOND 318

Orbital / Satellite Launch Market Share, Global = SpaceX Rising

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)

轨道 / 卫星发射市场份额，全球 =SpaceX 崛起

按年份和国家进行的轨道发射 – 1957-2025，每 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)

SpaceX Starlink @ 5MM+ Subscribers = +202% Annual Growth Over 3.2 Years

Starlink Global Number of Subscribers (MM) – 2021-2024, per SpaceX Announcements

Source: SpaceX public announcements

SpaceX 星链 @ 500 万 + 用户 =+202% 3.2 年年度增长率

星链全球用户数量（百万） – 2021-2024，根据 SpaceX 宣布

Source: SpaceX public announcements

SpaceX Starlink Ecosystem = Coverage Expanding Globally

SpaceX 星链 生态系统 = 覆盖范围正在全球扩展

Starlink Global Coverage – 5/25, per SpaceX

Source: SpaceX website (5/25)

星链 全球覆盖 – 5/25, per SpaceX

Source: SpaceX website (5/25)

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.

Source: SpaceX website (4/25)

Starlink =
Unlocking Previously-Inaccessible Internet Access in AI Era

选择全球 Starlink 应用案例 – 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.

Source: SpaceX website (4/25)

Outline

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

Outline

1. 似乎变化正在比以往任何时候都快？是的，就是这样
2. AI 用户 + 使用 + 资本支出增长 = 前所未有的
3. AI 模型计算成本高 / 上升 + 每令牌推理成本下降 = 性能趋同 + 开发者使用量上升
4. AI 使用 + 成本 + 损失增长 = 前所未有的
5. 人工智能变现威胁 = 竞争加剧 + 开源势头 + 中国崛起
6. AI 与物理世界加速 = 快速 + 数据驱动
7. 由人工智能驱动，从起步阶段就加速全球互联网用户增长 = 我们从未见过的增长
8. AI & 工作进化 = 真实的 + 快速的

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.*

AI 正从根本上改变我们的工作方式。随着物理自动化（想想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，其中 AI 系统可以进行推理、创造和解决问题。其影响范围广泛。

AI 认知能力的改进速度令人惊叹。自 ChatGPT 于 11/22 公开发布以来三年间，我们已从高中生的推理能力发展到博士生的水平。以大量结构化历史数据为输入并输出基于规则的决策和判断的职业，完全属于生成式 AI 的核心能力。在这个新兴的领域，劳动单位可能从人工小时转变为计算能力。数据中心和基础模型 – 在许多情况下 – 可能决定某些类型劳动的可用性和质量。

因此，一些人吹嘘一个 ' 代理未来 '，其中 AI 代理将取代许多白领工作中的人类。尽管可能，但历史和模式识别表明人类的作用是持久且引人入胜的。技术性飞跃通常推动生产力和效率的提升，并创造更多 – 但新的 – 工作。但话说回来，这次发生得更快。

在一个极端、完全自主的未来，人类在系统中保持一定的角色，转向监督、指导和培训。想象一下充满人类教导机器人复杂动作的设施，或者满是工人提供强化学习 * 人类反馈（ RLHF ）以优化算法的办公室。这不是推测。像 Physical Intelligence 和 Scale AI 这样的公司，分别基于对世界的这种看法建立了强大的商业帝国。

让人类劳动力重新配置以教导和改进机器作为主要功能的想法可能听起来是反乌托邦的。但值得记住历史上的相似之处。五十年前，这种成排的隔间和穿着制服的办公室工作者每天安静地坐在 LED 电脑前十小时的可能性，可能听起来同样反乌托邦。然而，我们就在这里。

技术不断重新定义和演变工作的性质和生产率 …… 人工智能也不例外。

**Reinforcement Learning = 一种机器学习方法，其中代理通过接收针对行为的奖励或惩罚来学习 .*

AI Impact on Business = Diverse & Broad

Industries That Could Be Affected by AI, per NVIDIA

Note: Global data shown. Source: NVIDIA

AI 对商业的影响 = 多样且广泛

可能受 AI 影响的行业，根据 NVIDIA

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 在劳动力中 – Shopify = 反射式 AI 使用现已成为基本预期 …

我们正进入一个时代，可以创造的商家和企业家可能比历史上任何时候都多。我们经常谈论降低复杂性曲线，让更多的人选择这个作为职业。创业道路上的每一步都充满了需要技能、判断力和知识的决策。在旅程中有 AI 陪伴，并且越来越多地不仅提供咨询，还为我们商家完成工作，这是一个令人震惊的阶跃式变化。

我们在 Shopify 的任务是让我们的软件毫无疑问是开发未来最佳企业的最佳画布。我们通过让每个人都保持前沿，并运用所有最好的工具，使我们的商家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 使用现在已成为基本预期。

也许你已经在那里，并发现这份备忘录令人困惑。如果是这样，你已经将 AI 作为思想伙伴、深度研究员、评论家、导师或配对程序员来使用。我一直在使用它，但我甚至觉得我只是触及了表面。这是我职业生涯中看到的最快的转变 ……

…… 有效地使用 AI 现在是 Shopify 每个人的基本预期。它今天是一种通用的工具，并且只会越来越重要。坦率地说，我认为不学习在您的专业领域应用 AI 的技能是不可行的；你可以尝试，但我必须诚实地告诉你，我认为今天不可能这样做，明天更不可能。停滞不前几乎是肯定的，而停滞不前是缓慢的运动失败。如果你不进步，你就会滑倒 ……

Shopify 联合创始人兼 CEO Tobias Lütke 在关于 AI 的内部备忘录中 – 3/25

来源：Tobi Lutke 通过 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 在劳动力市场 – Duolingo = Duolingo 将以 AI 为先

我在问答和许多会议中都说过，但我想使其正式：Duolingo 将以 AI- 为先。

AI 已经改变了工作的方式。这不是一个是否或何时的问题。它正在发生。当出现如此大的转变时，你能做的最糟糕的事情就是等待。2012 年，我们押注于移动。虽然其他人专注于为网站开发移动伴侣应用程序，但我们决定构建移动优先，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 年度应用程序，并解锁了随后发生的有机口碑增长 …

…AI 不仅仅是一种生产力提升。它帮助我们更接近我们的使命。为了教得好，我们需要创建大量的内容，而手动完成这些内容无法扩展。我们最近做出的最佳决策之一是用 AI 驱动的流程替换缓慢的手动内容创建流程。没有 AI，我们需要几十年才能将内容扩展到更多学习者。我们欠我们的学习者尽快提供这些内容 …

… 作为人工智能 - 优先意味着我们需要重新思考我们工作的许多方面。对为人类设计的系统进行微调将无法实现这一目标 …… 我们不能等到技术达到 100% 完美。我们宁愿紧急行动，并偶尔在质量上受到一些小打击，也不愿缓慢行动而错失良机。

我们将推出一些建设性的约束来指导这一转变 ……：

- …… 人工智能的使用将是我们招聘时考虑的一部分
- 人工智能的使用将是我们绩效评估时考虑的一部分
- 只有当团队无法自动化更多工作时，才会提供人员编制
- 大多数职能都将有具体的计划来根本性地改变它们的工作方式 ……

多邻国联合创始人兼 CEO 路易斯 · 冯 · 阿恩在关于 AI 的全员备忘录中 – 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：所问问题是"在过去六个月中，这家企业是否在商品或服务的生产中使用了人工智能（AI）？"BTOS 数据代表了美国经济中所有雇主企业，不包括农场。BTOS 样本由约 120 万家企业组成，进行双周数据收集。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，"2025 年第一季度：采用进展 modest，劳动力影响仍然微不足道"（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)

美国使用 AI 的企业比例变化 – 2024 年第四季度 -2025 年第一季度，根据美国人口普查局和高盛研究

注意：问题为“在过去六个月中，这家企业是否在商品或服务生产中使用人工智能（AI）？”BTOS 数据代表了美国经济中所有雇主企业，不包括农场。BTOS 样本包括约 120 万家企业，进行双周数据收集。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，“2025 年第一季度：采用进展适度，劳动力影响仍可忽略”（3 月 25 日）

AI Impact on Workforce = Employers Adopting AI to Drive Productivity Improvements

Objectives of Corporate AI / LLM Initiatives – Q3:23-Q3:24, per Morgan Stanley & AlphaWise

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

AI 对劳动力的影响 = 雇主采用 AI 以推动生产力改进

企业 AI / LLM 项目的目标 – Q3:23-Q3:24，根据摩根士丹利 & AlphaWise

来源：摩根士丹利，"GenAI：我们在哪里看到采用以及对 '25' 什么重要？（11/24）

AI Impact on Workforce = Seeing Productivity Gains, per Stanford HAI

Impacts of AI on Worker Productivity – 4/23, per Stanford HAI

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)

对劳动力的影响 = 看到生产力增长，根据斯坦福 HAI

对工人生产力的 AI 影响 – 4/23，根据斯坦福 HAI

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%

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%

注意：'AI 工作 ' 是指需要 AI 技能的工作。工作发布中 AI 技能要求使用马里兰大学的语言处理模型确定。- 仅限美国的工作大学 - 相关数据。数值已四舍五入。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

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

注意：本报告中的数据来源于 ZoomInfo 的专有专业联系人数据库 —— 一个每天检测超过 150 万人事变化的领先平台。为了编制职位名称的趋势，ZoomInfo 的数据科学家分析了数百家公司发布的关于他们从 2022 年 1 月 1 日到 2024 年 6 月 30 日的 AI 职位的公告。ZoomInfo 的数据库包括 1 亿家公司，34 亿名专业人士，& 1100 万名 C 级高管。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)

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.

就业演变 – 苹果 =600+ 生成式 AI 职位招聘

苹果与 "生成式 AI" 相关的职位发布 – 5/25

Source: Apple (4/25)

示例职位描述：

作为团队的一员，您将负责提出创新理念，并应用现代机器学习方法解决重要问题。从构思到产品化，您将参与核心技术开发的完整周期，包括手写和文本识别、手写合成、文档理解、自由形态绘图识别和生成。理想的候选人应具备计算机视觉、语音识别、深度学习和 / 或其他机器学习系统应用方面的经验。

USA Labor Productivity = Has Happened Alongside Job Growth Over Seventy-Seven Years

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)

USA Labor Productivity = Has Happened Alongside Job Growth Over Seventy-Seven Years

Relative Change in USA Non-Farm Employment & Labor Productivity – 1947-2024, per Federal Reserve Bank of St. Louis

注意：我们将个人电脑时代（PC Era）的起点定义为 1981 年（IBM PC 发布）。我们将桌面互联网时代的起点定义为 1995 年（Netscape 公司首次公开募股）。我们将移动互联网时代的起点定义为 2007 年（苹果公司 iPhone 发布）。我们将人工智能时代（AI Era）的起点定义为 2022 年（ChatGPT 公开发布）。资料来源：圣路易斯联邦储备银行（2024 年）（Netscape 公司首次公开募股）。

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 在劳动力中 – NVIDIA = 你不会失去 … 你的工作给一个 AI…*[但是]* 给一个使用 AI 的人

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…

… 但是让我给你提供两个你可能想要考虑的极端情况。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 万人知道如何编程和使用这项技术来达到极致 …

… 另外八亿、七亿半亿的人则没有。我会摆上台面的是，事实上，人工智能是我们弥合技术鸿沟的最大机会。让我来证明这一点给你看。你知道，如果我们只是看看这个房间，很可能超过一小部分人知道如何用 C++ 编程，而同样数量的人知道如何用 C 编程。然而，100% 的人都知道如何用 AI 编程。原因是因为 AI 会说任何你想说的语言 …

… 使用 ChatGPT 和 Gemini Pro 以及这些 AI 的数量在某种程度上表明，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 CEO 黄仁勋 @ 米尔肯全球会议 – 5/25

Source: Milken Institute (5/25)

Summary…

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…

Summary…

想象一下，如果暂时没有互联网，你接下来的日子会有多不同。现代生活的方方面面 – 我们如何工作、如何交流、如何治理，等等 – 很可能会完全颠覆。互联网已经渗透到生活的大大小小的方方面面，– 对许多人来说 – 很难想象一个没有互联网的世界。

在接下来的十年或二十年里，想象一个没有人工智能的世界可能会感觉一样。

人工智能正在以惊人的速度重塑现代格局。它始于研究，现已扩展到各行各业的新兴核心基础设施 – 从客户支持到软件开发、科学发现 教育，以及制造业。本文旨在描绘人工智能的扩展速度和广度，特别关注使用趋势、成本动态、基础设施建设和早期盈利模式。主线很清晰：人工智能正在加速，触及更多领域，并越来越深入地融入工作方式。

推动这一增长的是全球范围内易于使用的多模态人工智能工具（如 ChatGPT ）在普及的移动设备上的可用性，再加上推理成本的急剧下降和模型数量的爆炸式增长。现在，无论是封闭式还是开源工具，都越来越普及且功能越来越强大，使个人开发者、初创企业和企业能够轻松地进行实验和部署。与此同时，大型科技巨头正在将人工智能更深入地融入其产品 – 推出副驾驶、助手，甚至代理，重新定义用户与技术互动的方式。无论是嵌入在 SaaS 中的智能，还是消费应用中的代理工作流程，界面层都在实时重写。

在计算方面，投资持续大幅增长。主要云服务提供商、芯片制造商和超大规模企业的资本支出达到新高，这是为了在规模上实现实时、高容量的推理而进行的竞争所推动的。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。这种水平的资本支出是否能够持续还有待观察，但随着 AI 逐渐靠近边缘 – 在车辆、农场、实验室和家庭 – 数字基础设施和物理基础设施之间的界限 continues to blur.

全球在构建和部署前沿 AI 系统的竞争中，越来越受到美国和中国之间的战略竞争所定义。虽然美国公司目前在模型创新、定制硅和云规模部署方面领先，但中国正在快速推进开源开发、国家基础设施和国家支持的协调。两国都将 AI 不仅视为经济顺风，也视为地缘政治影响力的杠杆。这些竞争的 AI 生态系统正在加剧对主权、安全和速度的紧迫性 …

…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.

…Summary

…In 在这个环境中，创新不仅仅是一种商业优势；它是一种国家姿态 .

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，美中两国在国际影响力方面的这场竞争很可能将由最快的先行者赢得。因此，美国需要一个聪明的国际战略来迅速支持全球的美国人工智能 …… 中国明智地认识到，如果一个国家采用中国的 AI 平台，它很可能在未来继续依赖该平台。美国最好的回应不是抱怨竞争，而是确保我们率先赢得这场比赛。这将要求我们迅速有效地推广美国人工智能作为更优越的替代方案。这还需要美国盟友和朋友的支持和参与。

最后，人工智能正在改变我们与世界互动的方式。随着负担得起的卫星连接扩大对偏远和未充分服务地区的接入，下一波互联网用户可能会通过人工智能原生体验 – 跳过传统应用生态系统，直接进入对话式、多模态代理。

同样地，人工智能在工作场所的采用正在加速，并有可能塑造人们在工作生活中所花费的三分之一时间的方式。随着使用模式的发展和单位成本的下降，我们可能正见证一个早期阶段的互联网，其中智能是默认的界面 – 可访问的、情境化的，并且越来越个性化。

这一切都因信息流和资本的日益增长和透明度而加剧 – 以及武器化案例的不断增加。

这发生在全球大国更公开地主张专制与民主议程的时期。随着技术和地缘政治日益交织，不确定性正在上升。

有一件事是确定的 – 人工智能的游戏时间已经到来，而且只会越来越激烈 …… 以及 genie 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 是一家全球性科技投资公司，支持有远见的创始人贯穿其创新和成长的全生命周期。BOND 的创始合伙人已经投资了包括 Airbnb、AlphaSense、Applied Intuition、Canva、DocuSign、DoorDash、KoBold Metals、Meta（Facebook）、Instacart、Peloton、Plaid、Revolut、Slack、Spotify、Square、Stripe、Twitter、Uber 和 VAST Data 在内的行业先驱。

本文件，包括其中包含的信息，仅为信息提供目的编制，并不构成出售任何证券的要约或购买任何证券的要约邀请。此类要约或邀请只能根据与该证券相关的招股文件（如有）进行。

本文件依赖于来自广泛来源的数据 + 见解，包括公共 + 私营公司、市场研究公司 + 政府机构。在数据是公开的情况下，我们引用了具体来源；本文件也受到非公开信息 + 见解的影响。我们对本文件作出任何 + 所有明示或暗示的保证免责。

我们不评论数据来源的相对真实性或由此明示或暗示的不一致之处。任何文档内容均不应被视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）上发布本文件的更新、修订或澄清。

BOND 拥有或曾拥有本文件中提及的某些公司的重大股权头寸。

Data Provided Under License to BOND

yipitDATA SensorTower SimilarWeb