LEONIS CAPITAL

OpenAI

Building the “Everything Platform” in AI

Aug 21, 2025

Table of Contents

Acknowledgements 5

Executive Summary 6

Investment Thesis 7

- Bull Case 7
- Bear Case 7
- Technical Assessment 8
- Valuation Assessment 8

Company Overview: From Non-Profit Research Lab to the "Everything Platform" 9

Market Opportunity: Owning the Interface to the AI Economy 11

Products and Technology: Building the Operating System for the AI Era 13

- GPT Series 13
- Codex and GitHub Copilot 17
- DALL-E and Sora 18
- Voice and Speech 18
- Enterprise Solutions 19
- Hardware 19

Business Model: The Economics of an AGI Platform 21

- **Customer Segments & Revenue Streams** 21
  - Consumer Subscriptions 21
  - Enterprise Customers 23
  - Developer and API Revenue 24
  - Cross-Segment Network Effects 25
- **Partnership with Microsoft: Platform Alliance and Tension** 25
  - The Strategic Alliance (2019-2023) 25
  - Tensions and Strategic Drift (Late 2023-2024) 26
- **Cost Structure** 27
  - People and Talent 27
  - Compute Infrastructure 28
  - Spending Towards Platform Dominance 28

Competition: The AI Platform Race **31**

Major AI Research Labs **31**

Anthropic 31

Google 32

xAI 33

Meta 33

Open-Source Alternatives **34**

Hyperscalers & Enterprise Ecosystem Players **35**

Amazon 35

Team and Governance: Building a Team Around Mission AGI **36**

Investment Thesis: Valuing the Everything Platform **39**

Conclusion **43**

DISCLAIMERS

This document is provided for informational purposes only and should not be construed as investment advice, a recommendation, or an offer to buy or sell any security, instrument, or strategy. Nothing herein should be relied upon as financial, legal, accounting, or tax guidance.

The content reflects information available to Leonis Capital (together with its affiliates, “Leonis”) at the time of publication, including third-party sources believed to be reliable and the author’s own analysis. While efforts have been made to ensure accuracy and completeness, Leonis does not guarantee that the information is correct, complete, or current. All views and opinions are subject to change without notice, and neither Leonis nor the author(s) accepts liability for errors, omissions, or reliance placed on this material.

This document may include forward-looking statements, projections, or illustrative examples. Such statements are inherently uncertain, depend on assumptions, and are subject to factors beyond Leonis’ control. Actual results may differ materially from those anticipated. Past performance is not a reliable indicator of future returns, and investing always involves risk, including the potential loss of principal. Different types of investments involve varying degrees of risk, and there can be no assurance that the future performance of any specific investment, strategy, company, or product referenced, directly or indirectly, will be profitable, equal any indicated performance level, or be suitable for any particular portfolio.

Leonis Capital is an Exempt Reporting Adviser (“ERA”) under the Investment Advisers Act of 1940 and applicable state regulations. ERA status does not imply a particular level of skill or expertise, nor does it represent endorsement by any regulatory authority. This material is intended for a general audience and does not account for the objectives, financial situation, or constraints of any individual, nor does it establish an advisory relationship with Leonis.

Leonis, its affiliates, and its clients may hold, increase, reduce, or dispose of positions in companies or securities mentioned, independent of any views expressed herein.

AUTHORS



Jenny Xiao

Jenny Xiao is a Partner at Leonis Capital. She was previously a researcher at OpenAI and received her Ph.D. from Columbia University.



Jay Zhao

Jay Zhao is a Partner at Leonis Capital. He is a long-time VC with 3 IPOs and 8 unicorns in his portfolio.



Liang Wu

Liang Wu is a Research Affiliate and former Senior Researcher at Harvard Business School, where he focuses on the evolution of business models with emerging technologies.

Acknowledgements

This report would not have been possible without the contributions of many individuals who generously shared their insights, data, and expertise.

We're grateful to the Leonis research fellows who contributed analysis and research, especially Oliver Wang (Ph.D. candidate at MIT).

Thank you to our data partner, Sacra, who provided crucial data and editorial feedback on many drafts.

We appreciate the AI researchers and engineers who reviewed early drafts and provided technical feedback.

Executive Summary

Founded in 2015 as a non-profit research lab, OpenAI has emerged as one of the dominant players in the AI industry. The company gained widespread public attention with the release of ChatGPT in late 2022 and quickly scaled to 700 million weekly users worldwide, driving revenue from $3.7 billion in 2024 to a projected $12.7 billion in 2025. OpenAI is on track to become the world's most valuable private company as it targets a $500 billion valuation in an upcoming secondaries round.

OpenAI has positioned itself as **the "Everything Platform" in AI**, going beyond building an everything app to become the foundational layer upon which all AI-powered applications, services, and interactions are built. This ambitious vision originates from its belief that artificial general intelligence (AGI) will be the most transformative technology in human history. The company envisions its platform as the convergence point for specialized AI capabilities across every domain of human activity, positioning itself as the orchestrator of digital interactions and transactions.

While some view the company's August 2025 release of GPT-5 as evolutionary rather than revolutionary, it highlighted how the company's continued push reinforces its platform power, especially across its different user segments. OpenAI's platform strategy leverages the advantages of capital and scale that compound over time in the AI market, with its first-mover advantage and massive investments creating a flywheel effect that strengthens both technical and distribution moats.

However, OpenAI's platform ambitions bring fierce competition from hyperscalers and other AI labs. Google's search dominance, Microsoft's productivity suite franchise, and Meta's social networks all face disintermediation if OpenAI successfully becomes the primary platform layer for AI interactions. This creates a complex dynamic where some of OpenAI's largest investors and partners are simultaneously becoming its most formidable competitors. OpenAI's relationship with Microsoft exemplifies this tension. AI-native competitors like Anthropic and xAI, and open-source projects like DeepSeek and Qwen, threaten to commoditize AI models.

OpenAI now faces the defining challenge of executing platform development across multiple dimensions simultaneously: maintaining technical leadership, building enterprise relationships, growing developer ecosystems, and achieving infrastructure independence, while competing against some of the world's most resourced technology companies.

Investment Thesis

Bull Case

1. **Technical leadership with deep moats.** OpenAI has proven its ability to achieve and reclaim, if not always maintain, a technical edge over its rivals in key model capabilities like reasoning (o series and GPT-5). Its massive user base also creates powerful data network effects and powers its distribution advantages.
2. **Explosive and scalable revenue growth.** Two-year CAGR of ~182%: $3.7B (2024) → $12.7B (2025) → $29.4B (2026). The company has viral adoption with high-margin enterprise customers and has stated that 92% of Fortune 500 companies use its products.
3. **Unmatched consumer distribution and brand.** With 700 million weekly users, OpenAI dominates the consumer market. ChatGPT is becoming synonymous with AI, like Google did with search, all the while compounding on its first-mover advantage.
4. **Decisive capital advantage.** The company raised a $40 billion funding round, which represents the largest private technology company raise ever, to sustain R&D and infrastructure investment. The company raises funding at a velocity and scale that competitors struggle to match.

Bear Case

1. **Capital dependence with delayed profitability.** OpenAI does not expect to be cash flow positive until 2029. Sustaining its lead requires continuous multi-billion-dollar raises in a high-interest-rate and potentially cooling AI funding market. A capital drought could force R&D slowdowns and erode its competitive edge.
2. **Commoditization of core model capabilities.** As OpenAI's platform edge largely relies on its technical leadership, it faces the threat of model commoditization from alternatives like DeepSeek. Model providers may undercut each other in pricing, and cheap open-source models also force margin erosion and threaten OpenAI's revenue unless OpenAI can justify continuous value for the price difference.
3. **Microsoft dependence and strategic misalignment.** Heavy reliance on Microsoft for infrastructure, distribution, and revenue share leaves OpenAI exposed to partner misalignment, pricing disputes, or regulatory separation. Losing this anchor partner would weaken enterprise adoption and compute access.
4. **Regulatory & Governance Overhang.** OpenAI must complete for-profit conversion by the end of 2025 to retain $20B of its $40B raise. The company is also facing escalating AI regulation, potential Microsoft-related antitrust scrutiny, and IP/privacy lawsuits that could limit product scope and growth.

Technical Assessment

OpenAI's technical lead rests on three key pillars: 1) its massive, optimized infrastructure from its partnership with Microsoft; 2) its advanced training and fine-tuning methods from its elite research talent; and 3) ecosystem and data network effects from ChatGPT's first-mover advantage. However, as the AI arms race goes on, OpenAI will need to invest relentlessly in talent, model architectures, ecosystem lock-in, and secure cheaper or exclusive compute (potentially via proprietary hardware) to maintain its edge against Big Tech rivals like Google's Gemini and Meta's Llama and newer entrants like Anthropic and xAI.

Valuation Assessment

OpenAI's valuation multiples reflect investor expectations for a company positioned to grow quickly and capture significant value in the AI transformation. At a $500 billion valuation, OpenAI would trade at 39.4x multiple on forward 2025 revenue of $12.7 billion (projected). This premium is significantly higher than the multiples of more mature AI platform companies like Nvidia (23.4x), Meta (9.5x), and Google (6.0x). It suggests that investors are prepared to invest in OpenAI regardless of how steep the valuation is to avoid the risk of missing out. Assuming it continues to be valued at the same multiples, OpenAI's valuation would increase to $1.2 trillion in 2026 based on its projected revenue of $29.4 billion, which would add it to the shortlist of trillion-dollar companies. Should valuation multiples contract to 20x revenue due to competition and macro market conditions, OpenAI's valuation would still be at $588 billion by 2026.

Company Overview: From Non-Profit Research Lab to the “Everything Platform”

OpenAI was co-founded in 2015 by Sam Altman, Elon Musk, Ilya Sutskever, Greg Brockman, Trevor Blackwell, Vicki Cheung, Andrej Karpathy, Durk Kingma, John Schulman, Pamela Vagata, and Wojciech Zaremba as a nonprofit research organization. Its founding mission was to ensure that artificial general intelligence (AGI) benefits all of humanity, and the lab sought to openly share AI research.

However, in 2019, OpenAI transitioned to a “capped-profit” for-profit structure to attract capital, while putting the for-profit arm (OpenAI LP) under a nonprofit entity (OpenAI Inc) to maintain mission focus. Microsoft’s $1 billion investment at the time marked a pivotal shift, funding OpenAI’s computing needs and establishing a close-knit partnership. This structural evolution reflects that building platform-scale AI capabilities requires massive capital investment that non-profit structures cannot support. The hybrid structure attempted to balance mission alignment with the capital requirements of (and the expectations that come with) platform development, a tension that continues to influence the company’s strategic decisions.

OpenAI was the first company to invest heavily in scaling Transformer-based AI models, resulting in the success of models like GPT-2, GPT-3, and later its first application product ChatGPT. The company’s early focus on scaling general-purpose models rather than building specific applications proved prescient for platform development. By creating foundational capabilities that could power multiple use cases, OpenAI positioned itself to serve as infrastructure for the broader AI ecosystem rather than just competing in individual application markets.

The company has since consistently advanced the capabilities of foundation models while successfully commercializing them through products like ChatGPT, GPT-4o, DALL-E, and Sora. However, these products serve dual purposes in OpenAI’s platform strategy: they generate revenue and user engagement and act as entry points into the ecosystem. ChatGPT, for example, functions simultaneously as a consumer application and as a funnel into the underlying APIs that enterprises and developers can integrate.

Despite generating $3.7 billion in revenue in 2024, the company continues to operate at a significant loss of $5 billion in 2024, due to its massive investments in computing infrastructure, talent acquisition, and research capabilities. This economic profile mirrors other platform companies during their growth phases: Amazon operated at a loss for years while building e-commerce and cloud platform infrastructure, and Uber sustained massive losses while scaling its transportation platform. The **platform strategy** prioritizes market capture and ecosystem development over immediate profitability, betting that the platform’s network effects will generate superior long-term returns.

The company’s current strategic direction reflects its ambition towards platform development across multiple dimensions:

- **Technical Foundation:** Continued investment in foundational model capabilities that can serve as platform infrastructure for diverse use cases and customer segments.
- **Ecosystem Development:** Building developer tools, enterprise solutions, and partnership programs that enable third-party innovation while creating platform lock-in.
- **Infrastructure Control:** Investments in compute infrastructure, custom hardware, and operational capabilities that provide platform advantages and reduce dependence on external suppliers.
- **Multi-Modal Expansion:** Developing capabilities across text, image, video, and audio modalities to serve as a comprehensive platform for AI-powered applications rather than competing in individual media categories.
- **Distribution Strategy:** Building direct customer relationships through consumer applications while developing enterprise and developer channels that provide platform access across market segments.

The flipside of OpenAI's platform strategy is the risk of strategic overreach. Unlike Amazon, which layered new businesses onto its platform over time, OpenAI is attempting to excel simultaneously at every level of the stack and across every customer segment, from building "Stargate" supercomputing infrastructure to developing consumer devices to competing head-to-head in enterprise AI software, even partnering with Mattel to embed ChatGPT into Barbie dolls. OpenAI frames these moves as part of a unified platform vision. But the breadth also raises a core risk: **loss of focus.**

This evolution from nonprofit research lab to would-be "Everything Platform" underscores both the scale of the opportunity and the execution challenge. Success will hinge on whether OpenAI can preserve its current advantages while selectively adding capabilities, rather than stretching too thin in an intensifying competitive landscape.

Market Opportunity: Owning the Interface to the AI Economy

The AI market represents one of the most profound technological shifts since the advent of the internet. What we're witnessing today **mirrors the early days of internet adoption**, when every business eventually became an internet business, except AI's transformation is happening at an unprecedented pace. What makes AI's market dynamics particularly compelling is its simultaneous disruption of both enterprise and consumer markets, creating an unprecedented opportunity for **platform convergence**. Traditional tech companies had to choose between being enterprise platforms (like Salesforce) or consumer platforms (like Facebook). AI enables a single platform to serve both markets simultaneously**.**

In the enterprise segment, AI has evolved from experimental pilot projects to core productivity infrastructure. The enterprise AI market spent approximately $24 billion in 2024, with this figure forecasted to reach $155 billion by 2030. As AI capabilities mature and adoption deepens, we expect the monthly revenue per user to increase, with a subset of users becoming power users that could spend hundreds, if not thousands, of dollars per month. The figures today, coupled with the progress of AI, suggest considerable room for growth within existing market structures.

On the consumer side, AI adoption has shattered traditional technology adoption curves. It took the iPhone eight years to reach one billion users. ChatGPT achieved a similar scale in just two years. This acceleration reflects AI's unique value proposition: rather than requiring users to learn new interfaces or workflows, AI adapts to human communication patterns.

The global AI market is entering a phase of explosive growth. Some estimates show AI becoming a trillion-dollar market by 2027, with different segments growing at a CAGR between 40% to 55%. Others estimate it to become a $1.3 trillion market by 2032 with an overall CAGR of 42%. Regardless of which figures you look at, it seems inevitable that the AI market is the next (multi) trillion-dollar market and is growing at a rapid pace. These figures, however, only scratch the surface of AI's true market potential. Unlike previous technology waves that primarily enhanced existing processes, AI is fundamentally reimagining entire categories of work and changing how consumers engage with technology, creating an opportunity to capture value across every vertical and use case simultaneously.

To grasp the market opportunity, it's essential to understand how AI value is created across different layers of the technology stack and why controlling the platform layer is so powerful. At the foundation level, infrastructure and compute providers enable the massive computational requirements for training and running AI models. The model layer, where OpenAI competes most directly, focuses on developing the core intelligence capabilities that become platform services. Above this sits the application layer, where AI capabilities are integrated into user-facing products and services.

Looking ahead, several market segments promise particularly explosive growth as AI models mature. Financial services will account for as much as 20% of global AI spending through 2028, with spending on specific use cases like augmented claims processing in insurance seeing a five-year CAGR of 36%. According to another estimate, AI-driven drug discovery will

grow into a $35 billion market in its own right by 2032, with a CAGR of more than 100% between 2023 and 2032. Specialized applications in healthcare, finance, and manufacturing command higher pricing and stronger customer retention than general-purpose models, suggesting a natural evolution path for the industry.

We're entering an era where, much like the internet before it, AI capabilities will become embedded in virtually every software product and service. For OpenAI, this environment presents the opportunity to become the foundational platform that orchestrates the entire AI transformation. Rather than competing in individual AI markets, the platform approach allows them to enable and capture value from all AI markets simultaneously.

Products and Technology: Building the Operating System for the AI Era

To achieve its “Everything Platform” vision, OpenAI must build what amounts to an operating system for intelligence, one that orchestrates reasoning, creativity, and problem-solving across domains. This requires not only frontier models but a suite of capabilities that serve as infrastructure for any AI application. OpenAI’s product strategy reflects this scope: GPT models as core reasoning engines, specialized services like Codex and DALL-E, and even hardware initiatives.

GPT Series

OpenAI’s core products are its Generative Pre-trained Transformer (GPT) models, which form the foundation of the company’s AI ecosystem. These models function as fundamental platform services, the equivalent of compute, storage, and networking in cloud platforms like AWS. GPT-3, released in 2020, demonstrated the ability to generate fluent text and marked OpenAI’s first commercial API offering. In 2022, the company introduced an improved GPT-3.5 and GPT-4 in March 2023. GPT-4 significantly advanced the state-of-the-art with its ability to handle more complex prompts and even accept image inputs.

Model	Release Date	Technical Breakthroughs
GPT-1	June 2018	First large Transformer LM; introduced unsupervised pre-training + supervised fine-tuning paradigm.
GPT-2	Feb 2019	Demonstrated that scale improves fluency; staggered public rollout due to misuse concerns.
GPT-3	June 2020	Few-/zero-shot learning breakthrough; launched commercial API; basis for Codex and Copilot.
GPT-3.5	Mar 2022	Introduced RLHF tuning; “text-davinci-002/003”; backbone of ChatGPT v1; edit/insert API features.
GPT-4	Mar 2023	Multimodal (text+image); state-of-the-art on benchmarks; much longer context windows.
GPT-4 Turbo	Nov 2023	Lower-latency, lower-cost variant; retains GPT-4 capabilities with extended (128K) context window.
GPT-4o	May 2024	“Omni” model natively supports voice, vision, and text; best-in-class multilingual and audio benchmarks.

o1	Dec 2024	Reflective chain-of-thought: "thinks" before answering, improving multi-step reasoning in math, coding, and science over GPT-4o.
GPT-4.5	Feb 2025	Codenamed "Orion"; largest GPT series model yet; 21% better coding performance over GPT-4o and 27% over GPT-4.5 preview; 1 million-token context window.
GPT-4.1	Apr 2025	1 million-token context window; 21% coding uplift vs GPT-4o; improved instruction-following; available as GPT-4.1, Mini, and Nano variants.
o3	Apr 2025	Private chain-of-thought planning; SOTA on GPQA Diamond (87.7%), SWE-bench Verified (71.7%), Codeforces Elo 2727; full tool-use via function calling.
o4-mini	Apr 2025	Cost-efficient multimodal reasoning: text+vision chain-of-thought, 128K-token context; strictly better cost-performance than o3-mini on AIME, GPQA benchmarks.
GPT-5	Aug 2025	Improvements in reasoning, speed, and application-building capabilities; reduced hallucinations and the ability to solve more complex problems.

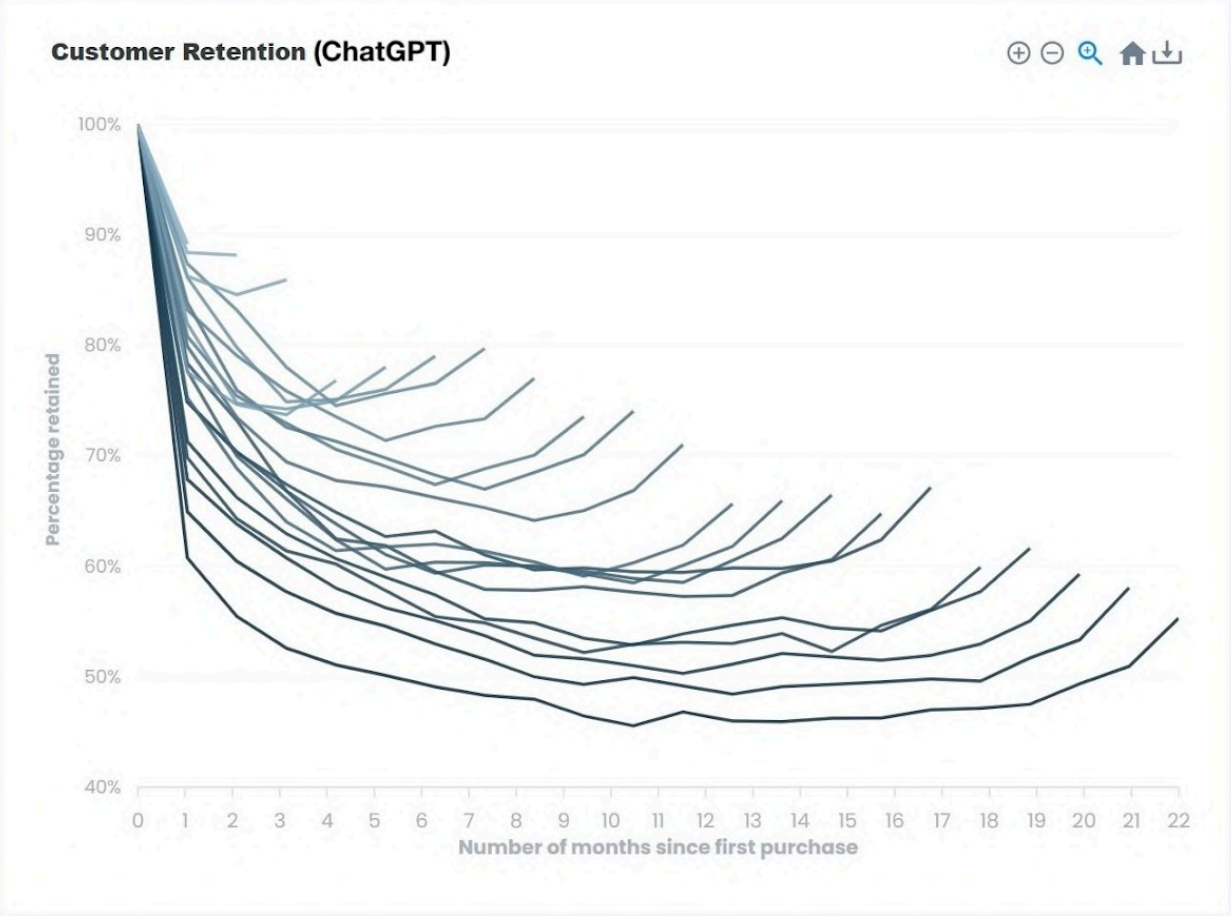
The application that catapulted OpenAI into public consciousness, ChatGPT, is a conversational AI chatbot based on GPT-3.5. Launched as a free research preview in November 2022, ChatGPT reached a million users in 5 days and 100 million in 2 months, becoming the fastest-growing consumer app ever. By early 2023, it was drawing mainstream attention for its ability to generate essays, code, and answers in a human-like dialogue format. OpenAI followed up with ChatGPT Plus (at a $20/month subscription) for enhanced capabilities and uptime, and by 2024, introduced ChatGPT Enterprise for business-grade security and performance. As of 2025, ChatGPT's usage is staggering; the latest data shows that the platform serves 700 million weekly active users, and independent analysis shows over 2.5 billion queries per day being handled.

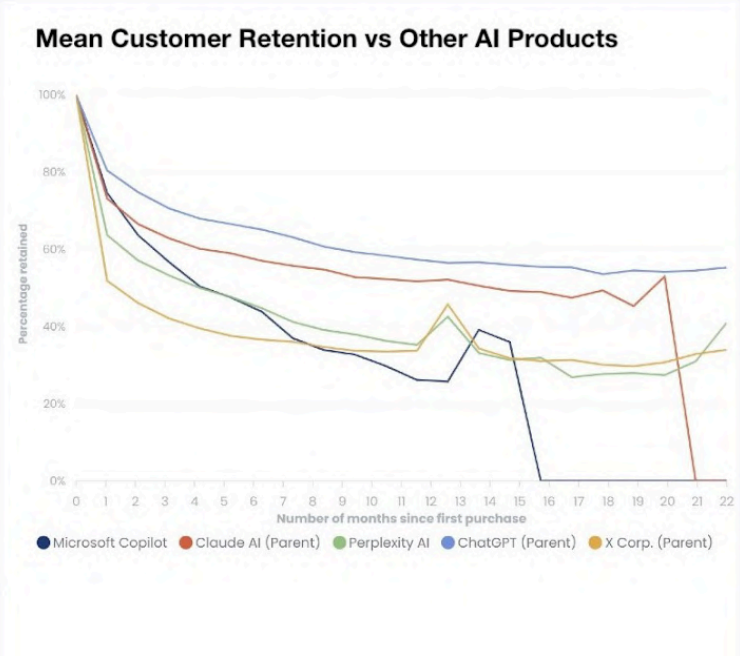
Following the launch of the multimodal 4o model, OpenAI quickly rolled out GPT-4.1, GPT-4.5, and the specialized "o-series." GPT-4.1 and 4.5 pushed improvements in coding, instruction following, and long-context reasoning, while models like o1, o3, and o4-mini targeted advanced scientific and coding tasks. The rapid cadence underscores a shift toward incremental feature iteration rather than breakthrough technical advances.

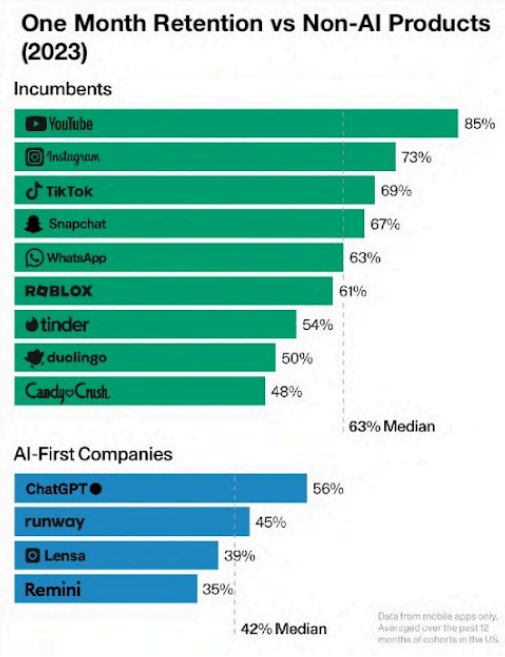
In July 2025, OpenAI also released its first open-weight models since GPT-2, the GPT-oss series, aimed at courting developers in response to competition from Meta and DeepSeek. This open-source model adopts a permissive Apache 2.0 license, allowing it to be used commercially without many of the restrictions in other models like Meta's Llama. Nevertheless, GPT-oss is more of a strategic hedge than a fundamental shift toward full openness.

The much-anticipated GPT-5 was finally released on August 7, 2025, after multiple delays from its original late 2024 target. Such delays stem from **fundamental scaling challenges**, bottlenecks in high-quality data, and potential limitations in GPU supplies due to Nvidia H100's production issues. Nevertheless, GPT-5 represents an improvement in reasoning, coding, and complex problem-solving capabilities. According to Cursor's CEO and some Cursor users, GPT-5 could surpass Claude 4.1 as the most intelligent coding model, which would have major implications for OpenAI's position in AI coding. Others have noted that GPT-5 was more evolutionary than revolutionary, with some users heavily critiquing the release. Some researchers have pointed out that other features like GPT-5's optimized routing could lay the groundwork for more sustainable monetization, even if technical gains felt incremental.

From a purely technical perspective, OpenAI's lead in LLM research is facing unprecedented pressure. Competitors like Anthropic's Claude 4, Google's Gemini 2.5 Pro, and xAI's Grok 4 have closed the capability gap significantly. But OpenAI's advantage increasingly lies in its platform strategy: ecosystem lock-in, distribution, and data advantage. Despite challenges in technical breakthroughs and unprecedented competitive pressure, ChatGPT maintains surprisingly robust user loyalty. The data shows ChatGPT's "smile-curve" recovery from initial churn to 50% to 60% retention in older cohorts, significantly outperforming other competitors. This retention strength stems partly from ChatGPT's early introduction of **memory and personalization features**, which create switching costs that pure capability comparisons miss.







Source: Deedy Das (Menlo Ventures). Data from mobile app only. More recent cohorts show a much higher early retention rate, indicating that the introduction of memory had an effect on improving retention.

This creates an interesting strategic tension: while OpenAI's technical moat narrows, its platform engagement moat may be widening, potentially allowing the **company to maintain market leadership even without clear technical superiority.**

On the core LLM technology front, OpenAI has several strategic paths forward. It can double down on breakthrough research for GPT-6, betting that a significant capability leap will reinforce its technical moat. It can also invest more heavily in multimodal and reasoning capabilities, where it's easier to achieve breakthroughs. But the focus on raw performance risks further delays while competitors close the gap. Alternatively, OpenAI can focus on vertical applications and enterprise integration, much like Anthropic's enterprise focus and its initial strategy to innovate on AI coding capabilities and now financial AI. This approach provides more predictable revenue but would put the company at risk of losing its edge in the core models if it underinvests. A third option involves doubling down on the user experience, personalization, and ecosystem features that deepen engagement rather than just improving raw capabilities. Of course, OpenAI will probably employ a combination of the above approaches, but the competitive AI race may ultimately force it to prioritize one over another.

Codex and GitHub Copilot

In mid-2021, OpenAI launched a specialized version of GPT-3 for writing code, called Codex, which powers GitHub Copilot. GitHub Copilot rapidly gained popularity among developers during its preview.[1] This period also saw other tech giants enter the fray (e.g., Amazon's CodeWhisperer and Google's AI code tools), but Copilot largely maintained its lead in developer mindshare.

However, by mid-2025, Codex and GitHub Copilot ceded significant mindshare to nimble competitors. While rivals like Cursor and Windsurf built AI-first IDEs with project-wide code intelligence, Copilot was stuck as a VS Code plugin and was limited by Microsoft's enterprise release cycles. OpenAI's bid to acquire Windsurf was an attempt to catch up on coding offerings, but the deal unraveled after it was vetoed by Microsoft.[2]

From a platform perspective, coding capabilities represent a critical service layer: better tools attract more developers, who in turn build more applications and drive greater usage and data for OpenAI. Yet in the coding assistants market, the company is at a crossroads. It can invest heavily to build its own IDE-integrated data pipelines, partner with third parties to license data similar to Windsurf's, or pursue smaller acquihires to replenish its talent and tech runway. Each path has trade-offs: in-house development avoids complex deal negotiations but takes time to match rivals' agility; licensing delivers instant data access at ongoing cost; acquihires can jumpstart capabilities but risk repeating Windsurf's integration pitfalls.

[1] However, there was some controversy over Codex regurgitating licensed code snippets.

[2] Under their partnership, Microsoft is entitled to all of OpenAI's IP until 2030 and refused to cede that privilege when it came to a new acquisition. This illustrates the challenges of building a platform while being dependent on partners who may have conflicting interests and can formally influence the outcome of critical strategic decisions.

DALL-E and Sora

Beyond text and code generation, OpenAI has established itself as a strong player in multimodal AI systems. DALL-E (2021) and DALL-E 2 (2022) introduced text-to-image generation, with DALL-E 2 garnering acclaim for its imaginative imagery and sparking broad interest in generative art. DALL-E 3 (2023) was integrated into ChatGPT and Bing, and in 2025, OpenAI embedded image generation directly into GPT-4o. The launch went viral with a "Ghibli-style" trend, highlighting how native multimodal tools can unlock consumer creativity and drive platform adoption.

The most ambitious expansion of OpenAI's multimodal capabilities is Sora, its text-to-video generation system. Released in February 2024, Sora can create high-quality video clips up to 20 seconds long from text descriptions.[3] Today, Sora includes sophisticated editing capabilities like remixing and storyboarding. However, OpenAI never established itself as the dominant player in image or video generation. Midjourney's artistic quality consistently outperforms the DALL-E series, while specialized competitors like Runway, Pika, and Kling boast better visual quality and editing features than Sora. Embedding DALL-E 3 and Sora into ChatGPT enables seamless use but limits optimization for professional creative workflows.

From a platform strategy perspective, multimodal capabilities represent service expansion that broadens the platform's addressable market. Rather than competing with specialized image and video generation tools head-on, these capabilities add to the list of reasons for users to visit its platform as a "one-stop shop." The more capabilities OpenAI bundles into its platform, the more jobs-to-be-done it can satisfy for the user. However, users who seek specialized services might be disappointed at the extent of OpenAI's offering in image and video generation. This reflects a strategic choice: OpenAI is prioritizing **a unified ChatGPT experience** over specialized tools, betting platform convenience will outweigh best-in-class performance.

Voice and Speech

Voice and speech expand OpenAI's multimodal interfaces, enabling natural conversation beyond text. Though still a small part of the product, voice is critical for ubiquity, ensuring access across modalities.

Its initial product was Whisper (2022), an open-source speech-to-text system that provides relatively accurate transcription and covers numerous languages. It served as the foundational infrastructure for voice-enabled applications built on the platform. OpenAI later introduced Advanced Voice Mode in September 2024.

The GPT-4o architecture breakthrough eliminated the traditional audio-to-text-to-audio pipeline, processing speech directly through a single neural network. But users report gaps between demos and production: it still behaves like transcription, voices sound robotic, and

[3] The system uses a diffusion transformer architecture that represents videos as collections of smaller units called patches, enabling training on varied visual data across different durations, resolutions, and aspect ratios. The company has also implemented safety measures, including C2PA metadata tagging and visible watermarks to enhance transparency and prevent misuse.

usage limits are low.[4] Meanwhile, specialist AI companies dominate the voice synthesis market. Rivals like ElevenLabs lead in professional voice synthesis with 3,000+ voice options and superior naturalness ratings.

Voice capabilities serve dual purposes for the platform: they expand the platform's accessibility and use cases while providing another interface layer that can capture user interactions and preferences. Voice is a great add-on to ChatGPT's platform experience, but OpenAI's approach suggests they view it as platform infrastructure rather than a standalone competitive product.

Enterprise Solutions

OpenAI has expanded into enterprise workflows with features like SOC 2 compliance, GDPR/HIPAA support, and data residency. Enterprise Agreements provide custom pricing and support, while ChatGPT Team and Enterprise let organizations build custom GPTs for functions such as coding, content, customer support, and HR.

Enterprise solutions represent a **critical platform access layer.** By providing enterprise-grade security, compliance, and administrative tools, OpenAI enables organizations to adopt the platform at scale while creating a pathway for platform expansion into regulated industries and large organizations that couldn't otherwise adopt AI capabilities. The partnership with Microsoft serves as a significant distribution advantage for OpenAI, making the company's models instantly available through the Azure OpenAI Service. But despite this advantage, the company faces fierce competition from hyperscalers who can leverage their existing enterprise relationships and integrated service offerings.

Having started in consumer/prosumer, OpenAI now faces a trade-off: double down on consumers, where it leads in distribution and UX but has worse unit economics and often requires subsidizing users; or strengthen its focus on the more profitable enterprise segment, where it must compete with its partner Microsoft, Google Cloud, and Anthropic. The challenge there is that OpenAI lacks the enterprise DNA and go-to-market strength of incumbents like Microsoft or Google.

Hardware

OpenAI is now expanding beyond software into hardware, representing a critical move for platform control. After reports in 2023 of Altman and Jony Ive exploring an AI device with SoftBank, OpenAI acquired Ive's design firm LoveFrom in 2025 for $6.5 billion to build a dedicated consumer product. Though success is uncertain, the move addresses a core vulnerability: distribution control. If AI assistants become as ubiquitous as smartphones or

[4] OpenAI also faced controversy around OpenAI's "Sky" voice, which sounded remarkably similar to Scarlett Johansson's voice from the movie "Her." This controversy illustrates the challenges of building platform services that must balance capability demonstration with legal and ethical constraints. After Johansson's legal team demanded transparency about how the voice was created, OpenAI quickly pulled the Sky voice from its platform.

search engines, OpenAI doesn’t want to be locked out by operating system owners like Google or Apple. By building its own hardware, OpenAI ensures direct access to users and control over the experience, much like how Apple’s iPhone became the distribution platform for mobile applications.

Business Model: The Economics of an AGI Platform

Unlike traditional software, where profitability comes from driving marginal costs toward zero, OpenAI **sells "intelligence by the token."** Each token represents a fraction of reasoning capacity that enterprises, developers, and consumers pay for. As usage grows, so do compute requirements and infrastructure expenses. The marginal costs don't go to zero like it does in traditional software businesses. To succeed, OpenAI must not only calibrate the right amount of intelligence per query to deliver value at a deflationary price point but also use R&D to discover more efficient architectures and serving techniques; otherwise, scale will amplify losses rather than create leverage.

At the same time, this model creates powerful moats for OpenAI. Competing against OpenAI requires billions in capital to replicate the infrastructure and ecosystem OpenAI has already built. The company's multi-sided platform strategy amplifies these dynamics by serving consumers, enterprises, and developers simultaneously, creating network effects where each segment's participation makes the platform more valuable for the other segments. For now, profitability takes a back seat to market capture, with OpenAI spending billions annually to entrench its position as a potential cornerstone of the next technology era.

Customer Segments & Revenue Streams

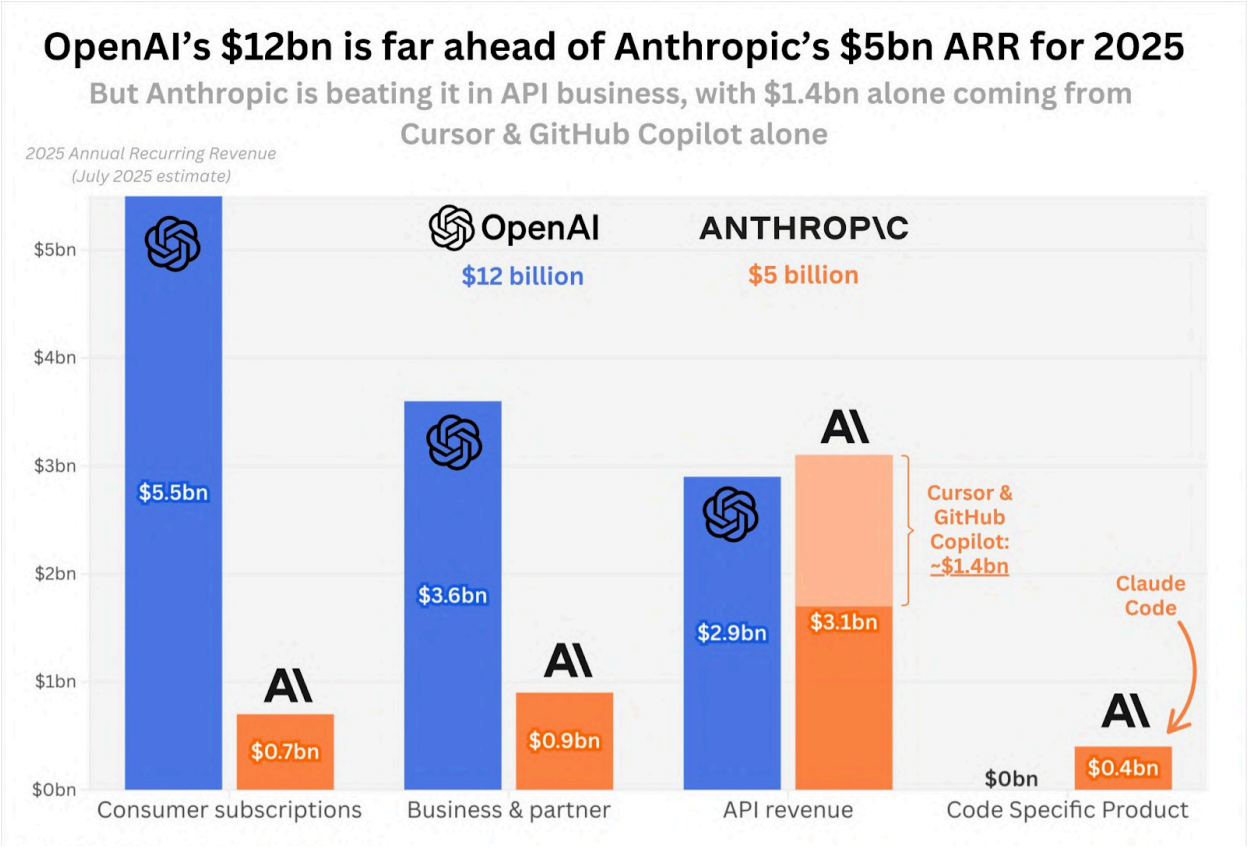
OpenAI generates revenue through three primary channels, each serving different customer segments with distinct pricing models and growth dynamics.

Consumer Subscriptions

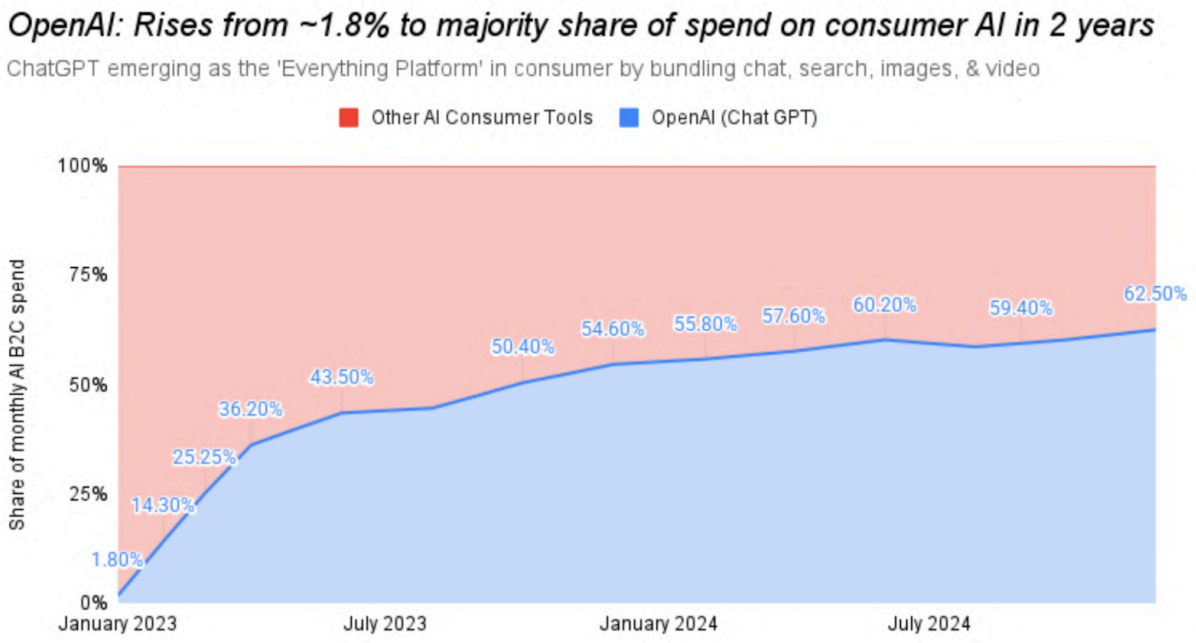
OpenAI's largest customer base consists of individual users accessing its AI capabilities through ChatGPT. As of April 2025, this segment has grown exponentially, with nearly 700 million weekly active users, a dramatic increase from the 400 million reported just two months earlier in February. The free tier continues to attract the majority of users, providing access to models like GPT-4o mini with some usage limitations. For premium features, individual users can choose between multiple subscription tiers: ChatGPT Plus ($20/month) provides access to GPT-4o and other advanced models with higher usage caps, while ChatGPT Pro ($200/month, launched December 2024) offers power users unlimited access to the most capable models and advanced features.

The subscription business has become a significant revenue driver, with ChatGPT Plus estimated to have around 20 million paying subscribers. ChatGPT Pro, its premium subscription, accounts for 5.8% of OpenAI's consumer sales by January 2025, with ChatGPT Plus representing the remaining 94.2%. These individual users span diverse demographics and professions, with particularly strong adoption among software developers (63%), marketing professionals (65%), and journalists (64%). While the United States remains OpenAI's largest market by user count, the service continues expanding across more than

160 countries worldwide. ChatGPT chief Nick Turley named India as a market OpenAI is particularly excited about for potential growth.



Source: Peter Gostev. Data from the Information.



Source: Earnest Analytics, Vela Gamma transaction data.

OpenAI's consumer segment operates as a strategic loss leader. CEO Sam Altman has acknowledged that even the $200 monthly subscription loses money due to high computational costs. However, this segment serves multiple strategic purposes: providing massive scale and training data, creating viral adoption that reduces customer acquisition costs, and showcasing capabilities to higher-value segments like enterprise customers. In short, the consumer segment operates at low or negative margins to drive network effects and ecosystem growth for the platform.

Despite subsidizing free users, OpenAI still projects approximately 50% margins for 2025. This figure is below Anthropic's 60% margins, which reflect Anthropic's heavier enterprise mix, and far lower than traditional software businesses. Nonetheless, even this margin level is notable given OpenAI's massive free-user base. Compared to other AI platforms that subsidize users even more heavily, such as Lovable (35% margins) and Replit (23% margins), OpenAI's margins are relatively strong, indicating that its consumer-first model could be more scalable than it appears.

Beyond subscriptions, OpenAI is exploring **advertising as a new revenue stream**. With its 700 million weekly users, even a modest uptake in advertising could rival early search in scale, and the conversational format provides a more natural entry point for commerce.[5] Sam Altman, once skeptical of ads, has floated lightweight referral and transaction models as potential venues for advertising.

At the same time, OpenAI is extending ChatGPT into **an identity layer** through its new "Sign in with ChatGPT" feature, which allows users to authenticate into third-party services with their OpenAI account and carry memory and personalization across apps. Taken together, consumer subscriptions may only be the first layer of a broader monetization stack, with ads and transactions positioning ChatGPT as something closer to a super-app than a standalone tool.

Enterprise Customers

The enterprise segment has emerged as a critical growth area for OpenAI, with offerings tailored to organization-wide AI deployment. ChatGPT Enterprise, launched in 2023, provides businesses with enhanced security, privacy controls, longer context windows, and administrative tools. This offering has gained remarkable traction, with OpenAI reporting 5 million paying business users, representing a more than 4x increase in enterprise customers since September 2024. The adoption spans companies of all sizes, though larger enterprises with over 10,000 employees show the highest utilization rates. According to OpenAI, 92% of Fortune 500 companies now use their technology in some capacity, reflecting the mainstream adoption of AI among major corporations.

Enterprises account for approximately 30% of OpenAI's total revenue ($3.1 billion from businesses and partnerships), and API services separately add another 12.5% ($2.9 billion). The enterprise adoption is driven by measurable productivity gains, with companies reporting

[5] In a recent piece, SemiAnalysis suggests that GPT-5's new "router" architecture can become a gateway for advertising. Low-value queries will be handled by cheap mini-models, while high-intent prompts such as product or service searches can be handled by agentic responses that include sponsored recommendations, affiliate links, or flat-rate referrals.

significant efficiency improvements: customer service productivity increases of 30-45%, marketing output boosts of 5-15%, and revenue growth of 3-15% for firms investing in AI. These concrete business outcomes validate the platform's value proposition and create case studies that drive further enterprise adoption.

ChatGPT Enterprise costs approximately $60 per user per month with 150+ user minimums and annual commitments. This is a 100% premium over Microsoft Copilot 365's $30 monthly rate and triple Amazon Q Business's $20 pricing. Despite higher costs, the product maintains strong adoption with over 1.2 million enterprise seats sold and 80% of Fortune 500 companies maintaining active accounts.

Beyond its core enterprise offerings, OpenAI is exploring new revenue streams. The company reportedly signs exclusive data-sharing and licensing agreements with enterprises in sectors like healthcare, law, and finance, allowing OpenAI to train on valuable proprietary data while creating specialized models for substantial fees. OpenAI is increasingly offering white-labeled AI solutions that power many customer-facing applications without visible OpenAI branding. Additionally, the company is making a push into consulting for high-value clients, including custom fine-tuning and specialized implementations. These deals both monetize valuable enterprise data and deepen product stickiness.

Developer and API Revenue

The third major customer segment for OpenAI consists of developers and organizations leveraging its models through API access. This channel allows developers to embed OpenAI's capabilities into their applications, products, and services. According to OpenAI, developer traffic has doubled in the six months leading up to February 2025, indicating accelerating adoption. The API offers access to the complete range of OpenAI models, including GPT-4.1, GPT-4o, DALL-E 3, Whisper (speech-to-text), and the o-series reasoning models, with tiered pricing based on model capability and usage volume.

OpenAI's API customers range from independent developers to major technology companies building AI-enhanced applications. The pricing follows a consumption-based model, with charges per token (for text) or per image, making it accessible to startups while scaling for enterprise needs. This pricing structure aligns platform success with developer success. As developer apps scale, OpenAI's usage and revenue increase proportionally. For higher-volume requirements, enterprises can access the models through Microsoft's Azure OpenAI Service, which offers additional enterprise features like regional deployment options, enhanced security controls,[6] and integration with Azure's broader cloud ecosystem.

The developer segment serves as OpenAI's innovation and expansion engine. Developers build applications for specialized use cases and niche markets that OpenAI doesn't address directly, expanding the platform's total addressable market (TAM). Their experimentation with novel use cases and integration patterns provides OpenAI with insights into emerging market opportunities and technical requirements. As developers build more sophisticated

[6] Enhanced security controls include private endpoints and a virtual network. An Azure Virtual Network is equivalent to a virtual private cloud (VPC) in AWS to isolate workloads from the public internet.

applications with OpenAI's models, some switching costs emerge, but compared to consumers and enterprises, it remains **the lowest of the three customer segments**.

Cross-Segment Network Effects

The three customer segments create reinforcing value through multiple interaction patterns:

- **Consumer-to-Enterprise:** Consumer adoption creates familiarity and demand among enterprise workers, making enterprise sales easier. Consumer use cases also provide proof points for enterprise buyers about the platform's capabilities and reliability.
- **Developer-to-Consumer:** Developer applications expand the platform's functionality for consumers, increasing engagement and retention. Breakout consumer apps showcase the platform's potential, attracting more developers.
- **Enterprise-to-Developer:** Enterprise use cases and requirements drive platform improvements that benefit all developers. Enterprise customers also serve as design partners for new features and capabilities.

All segments contribute to the data network effects, where usage data improves the underlying models, benefiting everyone on the platform. Consumer interactions provide broad usage patterns, enterprise usage provides domain-specific improvements, and developer feedback identifies technical gaps and opportunities.

Partnership with Microsoft: Platform Alliance and Tension

OpenAI's alliance with Microsoft is both its greatest enabler and its biggest constraint. Microsoft has funneled nearly $14 billion into the company, providing the capital, compute, and enterprise distribution that turned OpenAI into a global platform. At the same time, by locking in equity, revenue rights, and deep integration, Microsoft ensures privileged access to frontier AI while limiting OpenAI's independence. The partnership epitomizes the paradox when an incumbent backs a potential disruptor, fueling its rise while quietly containing the threat.

The Strategic Alliance (2019-2023)

Microsoft's initial investment was $1 billion in July 2019, which marked the start of an exclusive partnership aimed at developing advanced AI (ultimately AGI) on Microsoft Azure. In return, Microsoft secured an outsized share of the economic upside.

Microsoft is reportedly entitled to approximately 49% of the profits from OpenAI's for-profit entity, capped at a 10× return on its investment.

In addition to equity profit rights, the deal includes extensive revenue-sharing. Under the current agreement, Microsoft receives 20% of OpenAI's revenues up to $92 billion. Conversely, Microsoft pays OpenAI a share of revenues when it sells OpenAI-powered services (20% back for Azure OpenAI). The partnership also granted Microsoft broad rights to OpenAI's intellectual property through 2030. Microsoft can deploy OpenAI's models and research in its products with the assurance that OpenAI won't offer the same capabilities to a Microsoft competitor during the partnership term.

A cornerstone of the partnership has been Microsoft providing the massive cloud infrastructure needed to train and run OpenAI's AI models. Essentially, Microsoft became the extended infrastructure team for OpenAI. From a go-to-market perspective, Microsoft has been instrumental in distributing OpenAI's technology to end users and enterprises on a massive scale. The Azure OpenAI Service, in particular, has brought OpenAI's models into the enterprise mainstream. Microsoft's huge enterprise sales force and Azure customer base effectively became OpenAI's distribution network.

Tensions and Strategic Drift (Late 2023-2024)

Starting in late 2023, the OpenAI–Microsoft relationship began to experience tensions. During the November 2023 governance crisis, Microsoft CEO Satya Nadella took an unusually direct role by openly criticizing OpenAI's board for firing Sam Altman and even announcing he would hire Altman and key staff into Microsoft as a contingency. This intervention effectively forced OpenAI's board to reverse course and reinstate Altman within five days.

While the episode restored Altman's position, it also rattled Microsoft's confidence in OpenAI's governance and catalyzed a shift in strategy. In early 2024, Microsoft formed the Azure AI Foundry to host and support other AI models, including those from Google and xAI, on Azure. Microsoft even internally set a long-term goal to develop its own proprietary AI models that could rival OpenAI's. Microsoft identified OpenAI as a competitor for the first time in its 2024 annual report. The two companies, while partners, were now selling overlapping products. OpenAI's direct offering of ChatGPT Enterprise to businesses puts it in competition with Microsoft's Copilot products.

At the same time, OpenAI is moving towards greater autonomy and independence. Most notably, to avoid Azure lock-in, OpenAI started exploring multi-cloud partnerships.[7] In early 2025, it launched a project codenamed "Stargate" to build its own AI supercomputing centers.

[7] In July 2025, OpenAI announced its cloud partnership with Google Cloud.

By mid-2025, the partnership that once appeared seamless had clearly hit choppy waters. Several points stand out in the tensions:

- **Revenue sharing tensions:** OpenAI wants to reduce the hefty 20% share of revenue going to Microsoft, arguing it limits reinvestment and growth. It has proposed to give Microsoft more equity ownership in exchange for lowering or eliminating the revenue cut.
- **The "AGI Clause":** OpenAI's partnership agreement with Microsoft states that if OpenAI succeeds in creating AGI, then Microsoft's revenue share and IP rights will terminate.[8] Microsoft is pushing to remove or modify the AGI clause in current negotiations.
- **Veto rights:** Microsoft's agreements give it veto power in major decisions, including OpenAI's planned restructuring by the end of 2025, which it can leverage to negotiate more favorable terms.

In the medium-to-long term, the falling out of Microsoft and OpenAI seems inevitable given their clashing platform interests. The partnership could falter and come to resemble the Intel-Apple partnership post-M1, contractually intact but strategically irrelevant. For OpenAI, this means that their strategic priority will be to secure further capital and compute independence, as well as bolster their own enterprise capabilities, all essential elements for platform independence.

Cost Structure

OpenAI's ambitions are exceedingly resource-intensive. The company went from spending single-digit millions annually in its non-profit days to multiple billions by 2023-2025. Despite its revenue growth, the company continues to operate at a loss to the tune of $5 billion in 2024. The company does not expect to be cash flow positive until 2029. The bulk of costs is driven by talent and compute.

People and Talent

On the talent and personnel side, the company has expanded dramatically from around 770 employees in late 2023 to around 3,000 by mid-2025, an almost 4x increase in less than 18 months. This explosion reflects the transition from research lab to full-stack AI platform, driving a massive rise in payroll and R&D expenses.

OpenAI is organized into three teams focusing on different platform components:

- **Research Team:** Builds new models (GPT series, o-series), forming the platform foundation.

[8] The two companies have defined "AGI" in financial terms and will have achieved this when OpenAI's AI is capable of generating $100 billion in profits.

- **Applied Team:** Turns models into products like ChatGPT and the API, running training pipelines on tens of thousands of GPUs and managing inference for millions of users.
- **Deployment Team:** Adapts products for industries, manages enterprise integrations, ensures regulatory compliance, and feeds user insights back into research.

OpenAI has hired aggressively from Google Brain/DeepMind, offering substantial compensation packages. To counter competitive poaching, especially the recent moves by Meta, OpenAI has also issued multi-million dollar retention bonuses to nearly a third of its workforce. This underscores both the fragility of relying on a small pool of highly mobile researchers and the willingness to spend heavily to retain them. The broader war for talent shows no sign of slowing, and the question is whether OpenAI can sustain these costs while maintaining its position at the frontier or whether rivals with comparable resources will continue to erode its edge.

Compute Infrastructure

OpenAI's compute requirements are massive and growing. In 2020, Microsoft constructed one of the world's top five most powerful supercomputers exclusively for OpenAI's use, with over 285,000 CPU cores and 10,000 GPUs. For training GPT-4 and beyond, Microsoft has continued scaling Azure's GPU clusters. The Stargate program, launched with Oracle Cloud and SoftBank in 2025, adds 4.5 gigawatts of data center capacity supporting over 2 million AI chips, with total planned capacity reaching 10 GW in a $500 billion infrastructure program.

OpenAI has embraced alternative hardware. It has begun using AMD MI300 series accelerators via Azure, amidst an acute shortage and rising prices of Nvidia GPUs. The problem was so severe that Altman admitted GPT-4.5's rollout was slowed in early 2024 because they were "out of GPUs." OpenAI is also actively developing its first in-house AI chip, partnering with Broadcom and TSMC.[9] The strategy is to diversify supply chains, cut Nvidia dependence, gain leverage with vendors, and reduce costs at scale, without abandoning Nvidia.

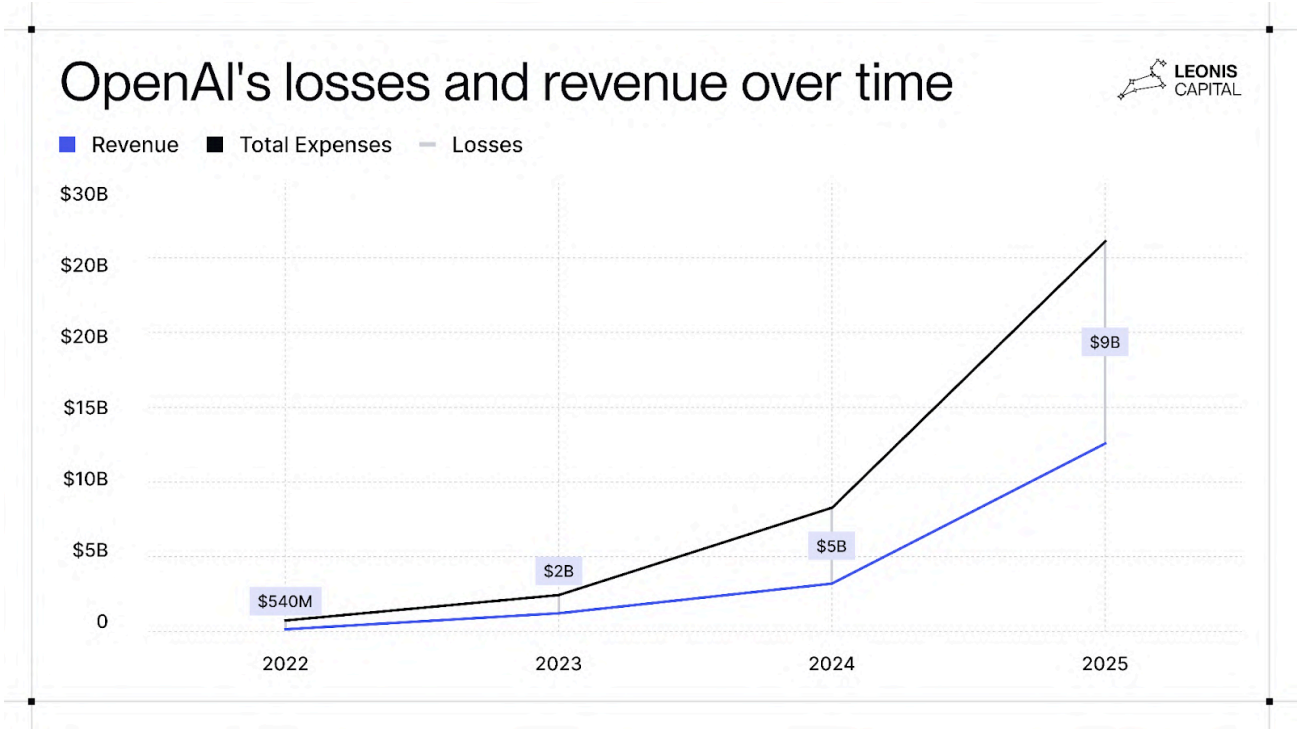
Sam Altman claimed that OpenAI is on track to have "well over 1 million GPUs" online by the end of 2025, dwarfing rivals like xAI, which reportedly controls approximately 230,000 H100s. Altman even quipped that the next hurdle is scaling 100x beyond that. Elon Musk has set a target of 50 million GPUs within five years, but even that would reach only half of OpenAI's projected capacity. A true 100 million GPU cluster would be aspirational, requiring an estimated $3 trillion in hardware alone. But the message is clear: OpenAI sees compute scale as one of the essential levers on the path to AGI.

Spending Towards Platform Dominance

[9] OpenAI has reportedly lined up TSMC's 3-nanometer process to manufacture it, targeting a first chip release by around 2026. This first chip will focus on AI inference rather than model training. The in-house silicon team has also grown from about 20 engineers in late 2024 to around 40 by early 2025, led by veterans of the Google TPU team.

Compared to competitors, OpenAI's spending is remarkable but not entirely alone in scale. Anthropic, for instance, will have raised a total of $25.7 billion after its most recent round, versus OpenAI's $63.9 billion total raise. Interestingly, their revenues line up in a similar proportion (OpenAI generating $12 billion in 2025 vs. Anthropic's $5 billion). The near one-to-one mapping between capital raised and revenue underscores how capital-intensive this sector is, where fundraising depth is a key factor that determines one's competitive position. Google DeepMind likely spends resources on par with OpenAI or greater. Elon Musk's xAI, with its recent $10 billion raise, is also pouring capital into compute.

A large part of building a sustainable business for OpenAI is finding the balance between its cost structure and revenue generation. Unlike traditional software platforms that can scale revenue without proportional cost increases, each API call, ChatGPT conversation, and enterprise query requires real-time inference on expensive GPU clusters. This creates a fundamentally different economic equation where traditional platforms achieve economies of scale, but **OpenAI's token economy requires physical infrastructure that scales with demand**.



Sources: 2022; 2023; 2024; 2025 revenue; 2025 losses.

While multi-billion-dollar capex has become normalized across frontier AI labs, a growing chorus of observers is alarmed by the unsustainable burn and its systemic consequences. OpenAI's reliance on steep losses, external subsidies (from Microsoft, Oracle, SoftBank), and fragile compute supply chains (e.g., GPU shortages, Stargate delays) creates precarious dependencies. Sequoia Capital's David Cahn called this AI's "$200B" and later "$600B" question that highlights the revenue gap in AI today. OpenAI's capital intensity could even create a systemic risk vector for the entire sector.

This deficit spending reflects OpenAI's strategy that betting on scale advantages will secure long-term market dominance. Each dollar invested in better models generates better performance, attracting more users and developers, generating more revenue and data,

enabling larger investments in the next generation. The scale required to compete means that only well-funded platforms can stay in the race.

Competition: The AI Platform Race

The battle for AI platform supremacy represents the highest-stakes technology competition since the early days of the internet, with trillion-dollar market opportunities hanging in the balance. Unlike previous platform wars that unfolded over years, the AI race is **compressing multiple competitive phases into simultaneous conflicts**: technical leadership, ecosystem development, enterprise adoption, and infrastructure control are all being contested at once. The stakes extend beyond AI itself. OpenAI's rapid growth threatens not just peers in model development, but the survival of traditional platforms whose core franchises (e.g., search, social media, enterprise service) risk being disintermediated by AI-native products.

Currently, OpenAI's defensibility rests on three main moats: (1) **frontier model leadership**, shipping cutting-edge reasoning and multimodal models faster than most competitors;[10] (2) **consumer scale**, with ChatGPT's user base giving it unmatched distribution and a real-time feedback loop; and (3) **enterprise integration**, anchored by its partnership with Microsoft, early-mover API adoption, and growing plugin/ecosystem layer. These moats are now under direct pressure in a platform-versus-platform battle for ecosystem control, rather than traditional product-versus-product competition. OpenAI's rivals, scaled AI research labs, open-source ecosystems, and hyperscale enterprise giants, are each attacking one or more of these moats in pursuit of their own platform dominance.

Major AI Research Labs

Several well-funded research labs are all pursuing platform strategies. These competitors represent the most direct threat to OpenAI's platform ambitions because they're building alternative foundations that could support competing ecosystems.

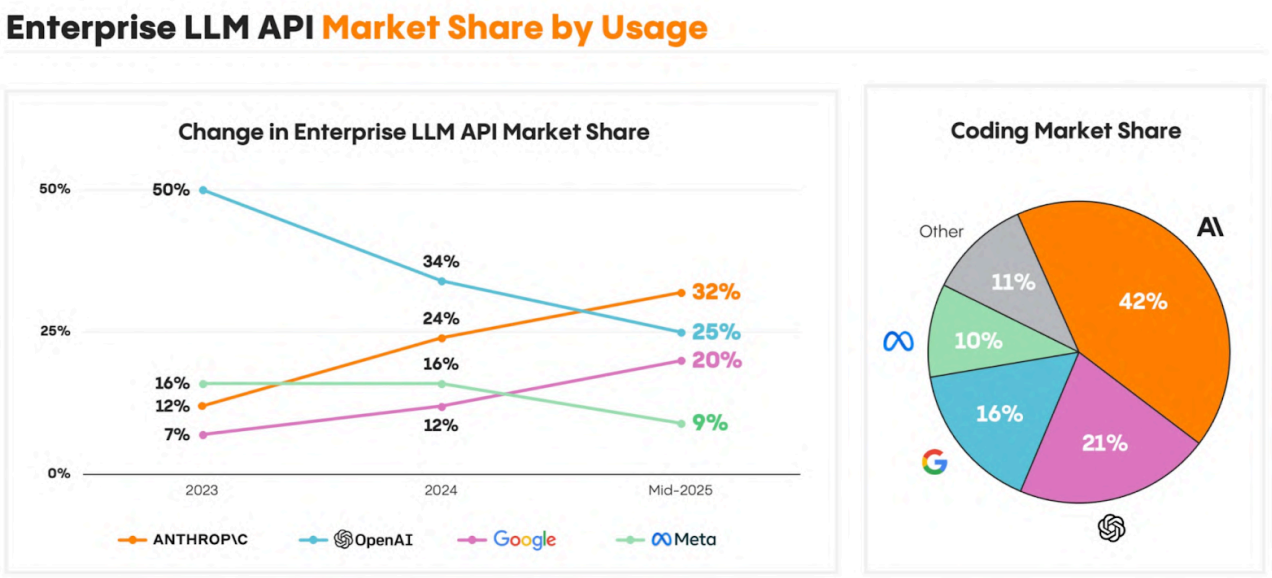
Anthropic

Anthropic, founded by former OpenAI researchers, has established itself as a formidable competitor with its Claude model family, which excels especially in coding and complex reasoning. The latest Claude 4 (in "Opus" and "Sonnet" variants) is regarded as the best coding assistant in the market.[11] Claude has also become the backbone of the 2025 "vibe coding" wave, powering AI-native IDEs and app builders like Cursor and Lovable, and serving as the system default in Vercel's v0 and Replit Agent. This positioning makes Anthropic not just a tool vendor but an ecosystem enabler, driving its API revenue to $3.1 billion, which accounts for more than half of its total ARR.

At the same time, Anthropic has leaned into an enterprise-first strategy, emphasizing safety and domain-specific solutions for highly-regulated industries like finance. By mid-2025, Anthropic's share of enterprise LLM usage had almost tripled from 12% to 32%, while OpenAI's share had fallen from 50% to 25%.

[10] Nick Turley, Head of ChatGPT at OpenAI, confirmed in a recent interview that ChatGPT works on a six-month roadmap, with longer horizons reserved only for enterprise commitments. This approach is highly unusual for a company of OpenAI's scale, as few other tech giant operates on such compressed timelines.

[11] Claude 4 scores 72.6% on a coding benchmark (SWE), far above GPT-4.1's 54.6%. OpenAI's GPT-5 (with thinking) claimed 74.9% accuracy on the SWE benchmark in its now infamous demo charts.



Sources: Menlo Ventures (2025).

Resource constraints remain Anthropic's weakness in the platform race; despite its recent $5 billion funding round, it operates with a fraction of OpenAI or Google's war chest. However, its focus on enterprise offerings and sectors like coding and finance means that Anthropic could build a dominant platform position in these high-value verticals. If the current trajectory continues, Anthropic could command half of the enterprise market by the end of 2025, potentially fragmenting OpenAI's platform ambitions.

Google

Google DeepMind represents the most formidable long-term platform threat because of its integrated ecosystem advantages. Gemini 2.5 Pro has demonstrated impressive capabilities in multimodal understanding and long-context performance.[12] Beyond raw performance, Google wields enormous platform advantages that OpenAI cannot replicate in proprietary data, custom TPU infrastructure providing 80% cost advantages over GPU-based solutions, and unparalleled distribution through Search, Chrome, Android devices, and Workspace tools. Google can deploy AI directly into products used by billions, creating immediate platform adoption without requiring separate customer acquisition.

Hyperscalers like Google are investing orders of magnitude more in capital expenditure than OpenAI. In 2025 alone, Amazon is spending around $100 billion, Google around $85 billion, Microsoft around $80 billion, and Meta $64–72 billion on AI-related infrastructure. While approximately 50% of that spend is on GPUs and compute infrastructure, the numbers dwarf what OpenAI can raise or deploy.

Google's Vertex AI has also rapidly gained ground and reached 15% market share by leveraging similar ecosystem advantages. Google can integrate AI capabilities directly with

[12] It scores 81.7% on the MMMU benchmark and maintains 83.1% performance at a 1-million token scale, far superior to OpenAI's best available long-context performance.

Google Workspace, BigQuery, and other enterprise services, creating compelling value propositions for organizations already using Google's enterprise stack.

Google's biggest challenge is execution speed and perception. Bard and Gemini's early stumbles hurt their reputation, and the current AI Mode on Google Search remains clunky compared to ChatGPT with its sleek UI. However, if Google successfully integrates advanced AI into its existing platform ecosystem, it could create a competitive moat that would be nearly impossible for OpenAI to overcome.

xAI

Elon Musk's xAI, with its Grok model series, represents another platform threat focused on real-time data integration. Grok's design reflects Musk's "edgy" ethos: it is styled to be witty and provocative, intentionally diverging from OpenAI's polished professionalism. Its latest models have demonstrated exceptional reasoning capabilities. Grok 4, released in July 2025, includes native tool use, real-time search, and outperforms rivals on several advanced benchmarks. Grok is also the only AI platform with real-time data integration with the X platform. This capability provides a significant advantage in applications requiring current information, such as journalism, market analysis, or social media monitoring.

However, Grok has had its share of public missteps. In a widely covered Kaggle chess tournament, OpenAI's o3 model swept Grok 4 (4–0), with commentators, including Magnus Carlsen, pointing out Grok's surprising blunders. More seriously, Grok 4 was jailbroken within days of release, and later generated antisemitic content and non-consensual deepfake nudes. These incidents highlight persistent safety weaknesses.

With access to substantial computing resources (xAI's 230,000 GPU Colossus cluster), Musk's backing, and the recent $10 billion raise, xAI represents a competitive threat focused on challenging what Musk perceives as OpenAI's deviation from its original mission.

Meta

Meta has emerged as a hybrid competitor: a trillion-dollar research lab with a deliberate open-source platform strategy.[13] Its Llama model family has become the foundation of the modern open-weight ecosystem, downloaded more than 350 million times by mid-2024 and adopted by enterprises like Zoom, Goldman Sachs, and Spotify for internal applications. While Meta lacks a flagship public chatbot, it has begun integrating generative AI across its existing platforms (Instagram AI stickers, WhatsApp chatbots, etc.).

Meta is investing unprecedented capital to close the gap with OpenAI. In 2025, Mark Zuckerberg announced hundreds of billions for new AI supercomputing centers (the

[13] While Meta often frames Llama as "open source," technically the models are better described as open-weights. The release terms fall short of OSI-approved open source licenses, most notably due to commercial use restrictions (e.g. the 700 million user cap) and broader licensing controls. This distinction has been a point of contention in the open-source community.

"Prometheus" 1 GW and "Hyperion" 5 GW datacenters), aimed at surpassing OpenAI's compute. Meta also reorganized its AI division to create the Superintelligence Labs after Llama 4's underperformance. They began poaching top talent, including OpenAI alumni, with multi-million-dollar offers.

Meta has long marketed open-source as a key differentiator, but that may be changing. In July 2025, Zuckerberg signaled that as Meta works toward superintelligence, it will "be rigorous about mitigating these risks and careful about what we choose to open source." Given the mounting infra costs and need to monetize R&D, Meta is shifting to a hybrid model, open-sourcing less powerful models while retaining exclusivity over breakthrough assets.

Open-Source Alternatives

The open-source AI ecosystem represents a fundamentally different competitive threat to OpenAI's platform strategy. Rather than competing directly for platform control, open-source alternatives threaten to commoditize AI capabilities, potentially undermining the platform economics that enable OpenAI to capture value from its ecosystem.

Open-source models like DeepSeek's R1, Qwen 2.5, and various Llama variants offer "good enough" performance at dramatically lower costs. DeepSeek's R1 model nearly matches OpenAI's o1 in artificial analysis quality while charging $2.19 per million tokens compared to OpenAI's $60 for o1, a difference of nearly 30 times. Even Google has now released its open-weight and highly efficient Gemma model, which has been downloaded more than 150 million times. Together, these efficiency gains strike at the core of OpenAI's platform economics.

What makes open-source alternatives particularly threatening to platform strategies is their ability to enable platform fragmentation. Rather than developers building on a single platform like OpenAI's APIs, they can deploy models locally or use multiple open-source alternatives, reducing platform lock-in and switching costs.

DeepSeek has rapidly gained recognition for its innovative model architectures and its "truly" open-source models.[14] What makes DeepSeek remarkable is its efficiency; the company achieved competitive results with significantly less computational resources than its Western counterparts by implementing a "mixture-of-experts" architecture that activates only relevant parts of the model for specific tasks.

Qwen 2.5, developed by Alibaba Cloud, performs comparably to OpenAI's GPT-4o in benchmarks like mathematical reasoning and coding, often surpassing it in specific tasks. But it offers significantly lower API costs ($0.38 per million tokens) compared to OpenAI's models ($1.25 per million input tokens).[15] This model is also under the Apache 2.0 license, enabling customization and local deployment, unlike OpenAI's proprietary models.

[14] Among state-of-the-art models, DeepSeek stands out as one of the few that are genuinely open-source. It provides the full model, detailed architecture, permissive licensing, and no catch.

[15] GPT-5 costs $1.25/million input tokens and $10/million output tokens.

These open alternatives generally lag behind frontier models in raw performance but offer compelling advantages in cost, privacy, and customization. Many organizations prefer them for sensitive applications where data sovereignty is paramount, or where the economics of proprietary APIs don't make sense for high-volume applications.

However, open-source alternatives face their own platform challenges: fragmentation, support burden, and the difficulty of coordinating ecosystem development across distributed contributors. While they can commoditize individual model capabilities, they struggle to provide the integrated platform experience that includes APIs, enterprise features, safety measures, and ecosystem support that OpenAI offers.

Hyperscalers & Enterprise Ecosystem Players

On the enterprise side, OpenAI faces its strongest competition from hyperscalers who can leverage existing ecosystem advantages to challenge OpenAI's platform ambitions.

Microsoft

Microsoft represents the most complex competitive threat because of its comprehensive enterprise ecosystem. Today, Microsoft maintains a 39% market share in foundation models through Azure OpenAI Service. Compared to Microsoft's deep ecosystem integration, OpenAI faces structural disadvantages in enterprise sales. Microsoft can bundle AI capabilities with existing Office 365, Azure, and enterprise software contracts, reducing friction and customer acquisition costs. For enterprise customers already embedded in Microsoft's ecosystem, choosing Microsoft's AI platform requires no additional vendor relationships or integration complexity. By contrast, OpenAI must sell as a standalone vendor, often encountering higher friction and longer sales cycles. Further complicating the picture is OpenAI's partnership with Microsoft (see callout section above).

Amazon

Amazon has entered the competition with its Amazon Bedrock service and Nova series of models, focusing on enterprise customers seeking managed AI infrastructure with strong security and compliance features. The company leverages its dominant cloud position to capture organizations looking for integrated AI solutions. Amazon's strategy is to make AI capabilities just another AWS service, deeply integrated with existing cloud infrastructure.

The competition OpenAI faces is increasingly characterized by platform vs. platform dynamics rather than traditional product competition. This competition is won on network effects, ecosystem lock-in, scaling infrastructure, and owning distribution channels. The ultimate platform competition question is not which company builds the best AI models, but which platform becomes the primary interface through which enterprises and consumers access AI capabilities.

Team and Governance: Building a Team Around Mission AGI

Building and executing an “Everything Platform” strategy requires organizational capabilities that extend far beyond AI research. Few technology companies have attempted to excel simultaneously at consumer products, enterprise sales, developer ecosystems, infrastructure operations, and strategic partnerships while maintaining technical leadership in a field evolving at breakneck speed. For OpenAI, this challenge is amplified by **governance instability and leadership turbulence**.

At the top, OpenAI is led by **Sam Altman (CEO)**, who guides overall strategy and serves as the public face. Altman is a co-founder and has been CEO since 2019 (briefly ousted and reinstated in November 2023). OpenAI’s leadership team has undergone significant changes since early 2025, reflecting the company’s evolution from a research-focused startup to a commercial AI powerhouse requiring diverse platform management capabilities.

The March 2025 leadership reshuffle included several key appointments designed to support platform execution across multiple business dimensions:



Greg Brockman – President & Co-Founder

Former CTO at Stripe and early OpenAI leader, now responsible for coordinating strategic execution across research and product units. His experience scaling payment infrastructure at Stripe gives him platform experience in managing complex, multi-sided business models.



Jakub Pachocki – Chief Scientist

Joined OpenAI in 2017 and led the development of GPT-4 and OpenAI Five. Promoted in May 2024 after Ilya Sutskever’s departure. Pachocki’s long tenure and technical leadership bring continuity in research direction while sustaining the pace of core model innovation.



Mark Chen - Chief Research Officer

Promoted from SVP of Research in March 2025 to lead OpenAI’s core model strategy and innovation pipeline. Chen’s role balances technical innovation with platform scaling, preserving leadership while expanding ecosystem reach.



Fidji Simo – CEO of Applications

Former CEO of Instacart and Meta executive, board member since March 2024. Officially assuming the CEO of Applications role in August 2025, overseeing product, business operations, and commercialization efforts. Simo's experience managing large-scale consumer platforms and marketplace dynamics directly applies to OpenAI's multi-sided platform strategy.



Brad Lightcap – Chief Operating Officer

Elevated in March 2025 to oversee operations, infrastructure, partnerships, and commercialization. Joined in 2018 after working with Altman at YC. Lightcap's expanded role reflects the complexity of operating global platform infrastructure across multiple customer segments.



Kevin Weil – Chief Product Officer

Joined in June 2024 to lead product strategy across consumer, enterprise, and developer offerings. Weil's experience scaling platforms at Twitter, Instagram, and Planet Labs equips him to align OpenAI's research advances with cohesive, user-driven products, which is key for expanding adoption and monetization.



Nick Turley – Head of ChatGPT

Previously helped launch ChatGPT from a hackathon sprint. He joined OpenAI in 2022 after leading product teams at Instacart. Turley embodies OpenAI's "maximally accelerated" product ethos (ship first, polish later) and prioritizes curiosity and hard questions over formal credentials.

Despite this strengthened leadership bench, OpenAI's trajectory has been marred by **governance instability and waves of high-profile departures** that create risks for platform execution. The first public sign of OpenAI's governance crisis came in November 2023, when OpenAI's board unexpectedly fired CEO Sam Altman. In May 2024, Ilya Sutskever and Jan Leike, co-leads of the Superalignment team, departed amid concerns that OpenAI's leadership had deprioritized safety in favor of product development. Several other top safety researchers left in the wake of the team's collapse. A few months later, co-founders John Schulman and Andrej Karpathy, as well as long-time CTO Mira Murati, also departed. Taken together, these departures reflect not routine turnover but ideological fracture and erosion of internal trust.

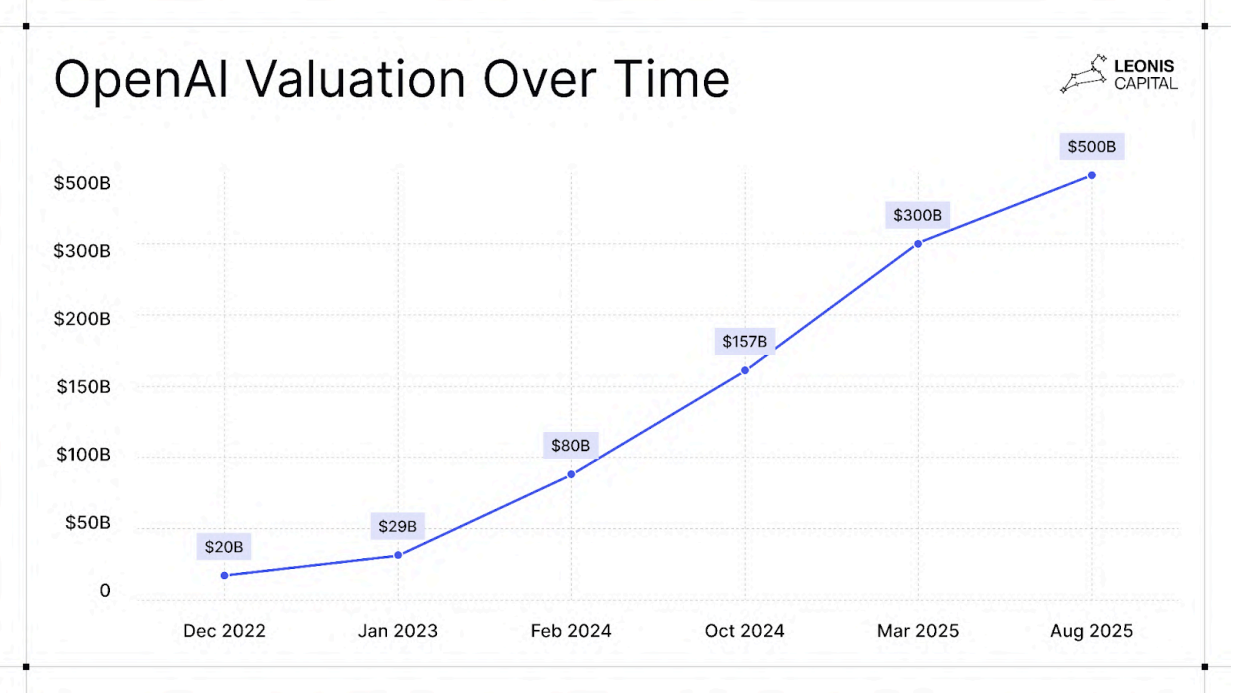
This leadership turbulence has been compounded by **broader attrition across the research ranks**. PitchBook reports that OpenAI has lost more than 50 senior researchers and over a quarter of its leadership within two years, with some leaving to found startups and others joining competitors. Meta has been particularly aggressive, hiring away several of OpenAI's leading scientists, including four directly responsible for GPT-4 and post-training on newer models. While Sam Altman has publicly minimized the impact, the cumulative departures point to a structural challenge of retaining elite researchers in a market where multi-million-dollar packages and even rumored $100 million signing bonus offers have become tools of competition.

Against this backdrop, funding terms now add further pressure. OpenAI must complete its for-profit transition by the end of 2025 or risk losing half of the $40 billion funding round. Elon Musk also filed a lawsuit challenging OpenAI's transition toward a for-profit structure, alleging it betrays the organization's nonprofit mission and potentially violates foundational agreements. This legal action introduces another roadblock to governance and structural transition. For OpenAI, **governance stability is as critical as technical leadership.**

Investment Thesis: Valuing the Everything Platform

Valuing OpenAI requires answering a complex question: **What will a company be worth if it successfully becomes the foundational platform for AI?** Traditional valuation methods struggle when applied to a company that could potentially capture value from every AI-powered application, service, and interaction across the global economy.

OpenAI's rise from $30 billion to $500 billion in just two years reflects investors' belief that they are not simply buying shares in an AI company, but acquiring stakes in what could become the most valuable platform in technology. The March 2025 SoftBank round wasn't just another funding milestone; it was an indication that the company was being underwritten as *the* operating system of the AI era. Roughly $18 billion of the $40 billion is earmarked for "Stargate," a $500 billion data center project aimed at building the compute backbone for global AI workloads.



Major investors include SoftBank, Microsoft, Thrive Capital, and Khosla, all of whom made fortunes building or backing platform plays. SoftBank's Masayoshi Son, in particular, has been a key figure in betting on platform shifts from the internet to the mobile era. He framed AI as a "defining force shaping humanity's future." This conviction has driven investors to price the company at a valuation usually reserved for entrenched category winners that dominate entire markets, not a startup that only recently rose to fame. This leads to the critical question: How do you justify paying $500 billion for such a young company?

The answer lies in how investors value *platforms*, not just products. Unlike products, platforms generate durable network effects, control critical infrastructure, and embed themselves into the workflows of millions, qualities that can compound into trillion-dollar market caps.

Leonis Capital Scoring: A Platform Valuation Lens on OpenAI vs. Big Tech

	OpenAI	Google	Amazon/ AWS	Apple	Meta	Nvidia
Distribution Advantage	4	5	4	5	5	4
Network Effects	4	4	4	5	5	4
Switching Costs	3	4	5	5	3	5
Technical Leadership	4	3	3	3	3	5
Scale Economics	2	5	5	4	4	4
Ecosystem Strength	3	5	5	5	4	4
Multi-Sided Markets	4	5	5	5	4	3
Overall Average	3.4	4.4	4.4	4.6	4.0	4.1
Financial Context (2025)						
Revenue	$12.7B	$400.0B	$710.0B	$417.0B	$200.0B	$192.4B
Valuation / Market Cap	$500B	$2.4T	$2.4T	$3.4T	$1.9T	$4.5T
Revenue Growth (YoY)	243%	14%	11%	7%	20%	48%
Forward P/S Multiple	39.4x	6.0x	3.4x	8.2x	9.5x	23.4x

Sources: “Financial Context (2025)” is based on data from Yahoo Finance; NASDAQ (August 8, 2025). Scoring from Leonis Capital research on a 1-5 scale.

To understand how investors are applying this lens to OpenAI, it helps to compare it against the most successful platform companies of the last two decades: Google, Amazon/AWS, Apple, Meta, and Nvidia. We believe there are **seven key factors that define platform leadership**: distribution advantage, network effects, switching costs, technical leadership, scale economics, ecosystem strength, and multi-sided markets. All of the platform companies

built a core technology layer that others depend on, but their dominance rests on different edges: Google's unmatched ecosystem strength, Amazon's scale economics in cloud, Apple's hardware-software integration, Meta's network effects in social platforms, and Nvidia's technical leadership in AI compute.[16]

By comparison, OpenAI is still catching up on several of these fronts. Its distribution advantage is meaningful, thanks to ChatGPT's reach, but nowhere near Meta's app user base or Google's default search positioning. Its ecosystem is growing but shallow, with limited third-party integrations compared to the app stores, cloud marketplaces, or partner networks that mature platforms command. Perhaps the most powerful edge for AI platforms is data network effects, where usage generates more data that can be used to train better models, improving performance and attracting more users. This is similar to the classic user-to-user or buyer-seller network effects of past internet, software, and marketplace leaders, and while OpenAI benefits from it, it's still early in turning that data advantage into an insurmountable moat and sustainable business.

Today, **OpenAI's biggest structural weaknesses lie in switching costs and scale economics**. Stickiness has improved recently: The "smile curve" in retention and the launch of memory and personalization features show progress, but the gains are uneven. For API users, switching costs are still negligible as developers can test or migrate workloads to Claude, Gemini, or open-source models with little friction. For ChatGPT and enterprise products, OpenAI is only beginning to build defensibility through personalization and ecosystem features, but these efforts are still early. Meanwhile, unlike most platform leaders, the LLM business model doesn't enjoy strong economies of scale. In fact, scaling is a structural challenge, as the costs for training and serving models grow disproportionately with usage. Overall, on almost every one of the classic success factors that define platform leadership, OpenAI is still an emerging player.

What OpenAI lacks in entrenched moats, it makes up for in velocity. History shows that investors have rewarded hyper-growth platform companies with extreme revenue multiples, well above today's incumbents. Nvidia, during the 2023-2024 AI boom, saw forward P/S peak around 25x, with trailing P/S climbing north of 40x. Google, post-IPO (2005-2007), commanded a P/E averaging 28x over a decade and spiked into the 50s on occasion. Meta in the early mobile shift (2012-2014) sustained P/S multiples in the high teens. These peaks coincided with periods when revenue was compounding 50–100% annually and investors prioritized velocity over durability.

By comparison, OpenAI's 39.4x multiple is higher than what any of these companies sustained, but justified by an even steeper revenue curve of 243% projected growth in 2025. The premium reflects not entrenched moats but market consensus that OpenAI's current growth resembles, in condensed form, the hyper-growth phases of past platform leaders. Public and private market heuristics often link valuation multiples to not only revenue but also the velocity of growth. The ERG (Enterprise Value/Revenue ÷ Growth) is a rule of thumb indicating that revenue multiples should be roughly a third of the growth rate. By that logic,

[16] The scores we offer here are intended to illustrate the relative strengths and weaknesses of these different platforms and not meant to be definitive measurements. They provide a rough comparison to contextualize OpenAI's positioning against more established leaders. Specific assessments may vary depending on methodology and interpretation.

OpenAI's growth rate could theoretically justify a multiple north of 60x. OpenAI's 39.4x multiple sits below that ceiling but far above the single-digit multiples of slower-growth incumbents, reflecting both a discount for competitive and scaling risks and a premium for market-shaping potential.

If we extend this math forward, the math gets even more striking. On projected 2026 revenue of $29.4B, OpenAI could command a valuation of $1.2T at current multiples. Even if competition drives multiple compression to 25x, which is roughly in line with Nvidia's current forward multiple, the company would still be worth $735B, almost 50% higher than today's valuation. But this also underscores how sensitive OpenAI's valuation is to sustained top-line expansion.

Investors are also betting that AI will become a bigger market than search, social, smartphones, or even cloud. Unlike these past platform waves, which were powerful but bounded categories, AI is a general-purpose technology that's closer to the internet in scope. Its surface area touches every industry: productivity, entertainment, healthcare, scientific discovery, and defense. In theory, **the TAM is the economy itself**.

This TAM logic explains unusual investor tolerance for capital intensity. OpenAI is projected to post losses of $5 billion in 2024 and $9-$14 billion in 2025, with similar or even growing burn rates likely for several years. However, OpenAI's backers are underwriting these numbers as table stakes for building the AI era's core infrastructure, much as platforms like Amazon, Tesla, and Uber absorbed years of heavy losses in pursuit of dominance. In platform plays, sustained negative cash flow is often seen as a feature, not a bug, when it fuels scale and lock-in.

Despite that platform opportunity, several factors could put pressure on OpenAI's valuation trajectory. Technical commoditization is a constant risk. In particular, if AI capabilities become undifferentiated faster than platform advantages develop, the valuation premium based on platform capture could collapse. Open-source alternatives and competitive convergence remain potent threats. Platform competition is also intensifying: Google's integrated ecosystem, Microsoft's enterprise distribution, and Meta's open-source strategy could fragment the AI platform market and shrink OpenAI's effective TAM. Execution risk looms large, as building an "Everything Platform" requires flawless coordination across technical, commercial, and partnership fronts, while mounting regulatory scrutiny on safety, competition, and societal impact could constrain growth or force structural shifts.

Against this backdrop, OpenAI's valuation is less a reflection of current revenue or any single product line, and more of a bet that it can close the moat gap with incumbents while AI's economic footprint meets optimistic projections. The premium multiple is sustainable only if the company achieves durable platform advantages that enable long-term value capture across the AI ecosystem.

Conclusion

When OpenAI was founded as a non-profit research lab in 2015, few could have envisioned its transformation into a company pursuing one of the most ambitious platform strategies in technology history. The question that has driven this entire analysis is whether OpenAI will succeed in becoming the “Everything Platform” it aspires to become.

The parallels to the early internet are striking. In the internet era, companies that controlled internet platforms, Google with search, Amazon with e-commerce, and Facebook with social media, captured extraordinary value by becoming essential intermediaries of digital interactions. Like the internet giants before it, OpenAI seeks to intermediate digital interactions, but on a far larger canvas: intelligence itself.

Its opportunity is massive: a $29.4 billion revenue trajectory by 2026, 700 million weekly users, and the strongest capital base in AI history. Yet its risks are equally stark: dependence on Microsoft and other key partners, exposure to commoditization, high burn rates, and intensifying hyperscaler competition. The “Everything Platform” vision forces OpenAI to execute on all fronts at once: technical leadership, enterprise lock-in, developer ecosystems, and compute independence.

The company’s march towards “Everything Platform” highlights the complexity of building comprehensive technology ecosystems in an era of accelerated competition, where the very speed required to capture platform advantages risks undermining the methodical construction that makes platforms durable. The company has proven that massive scale and rapid iteration can create market dominance, but the ultimate test lies in whether such velocity can translate into compounding moats that define lasting platforms.