

[← ALL PERSPECTIVES](#)

PERSPECTIVE • PORTFOLIO

# 2025 Mid-Year LLM Market Update: Foundation Model Landscape + Economics

A new enterprise LLM leader has emerged as usage and spend surge.

July 31, 2025

[Tim Tully](#), [Joff Redfern](#), [Deedy Das](#), [Derek Xiao](#)

## Anthropic Surpasses OpenAI in Enterprise Usage

---

## Open-Source Adoption in the Enterprise Flattens

---

## Enterprises Switch Models for Performance, Not Price

---

## AI Spend Is Moving from Training to Inference

---

By using this website, you agree to our use of cookies. We use cookies to provide you with a great experience and to help our website run effectively.

[Accept](#)[Decline](#)



When we released Menlo Ventures' [2024: The State of Generative AI in the Enterprise](#) report last November, several critical questions about this foundational layer remained unanswered:

Would demand for LLM APIs keep pace with the growth of consumer applications?

How smart will these models get, and how fast will they get there?

Would open-source models catch up to closed-source frontier models in performance, and if so, how would that impact enterprise adoption?

And most importantly, where might long-term value accrue?

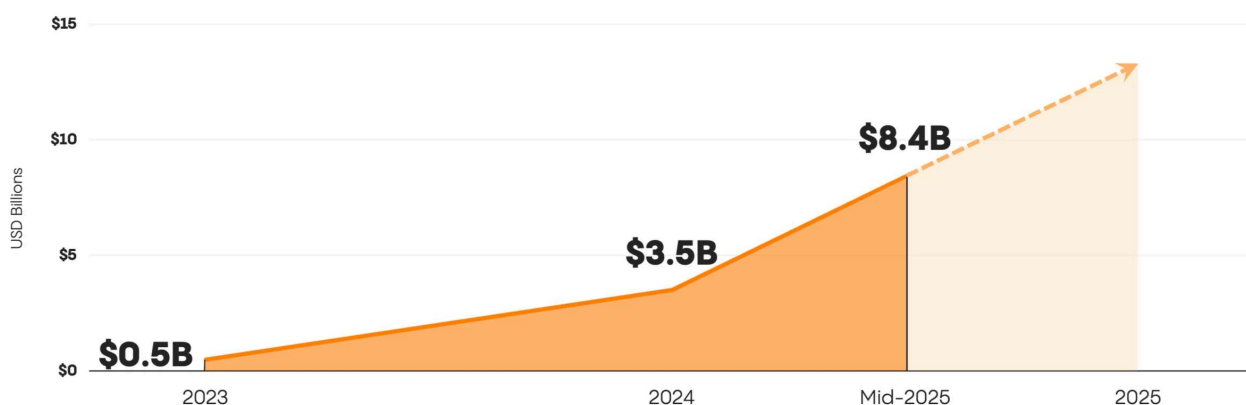
Six months later, the data tells a clearer story:

Model API spending has more than doubled in this brief period—jumping from **\$3.5 billion** (of a total **\$13.8 billion** generative AI spend we estimated last year) to **\$8.4 billion**.<sup>1</sup> Enterprises are increasing production inference rather than just model development, marking a shift from previous years.

Code generation has become AI's first breakout use case. Beyond pre-training, foundation models are now scaling along a second axis: reinforcement learning with verifiers. And while open-source continues to advance, the slowdown in frontier breakthroughs from Western labs has tempered what had previously been a rise in enterprise adoption. As a result, enterprise dollars are now consolidating around a few high-performing, closed-source models, giving us a new market leader in [Anthropic](#)\*.

To capture the state of the current LLM market, we surveyed over 150 technical leaders<sup>2</sup> across startups and enterprises on the modern AI stack's foundation layer: who's gaining share, what's running in production, and the selection criteria shaping the entire stack.

By using this website, you agree to our use of cookies. We use cookies to provide you with a great experience and to help our website run effectively.



© 2025 Menlo Ventures

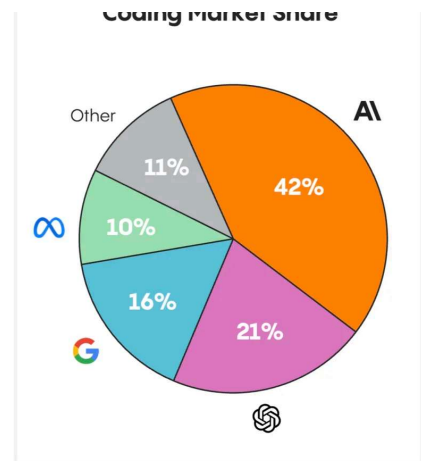
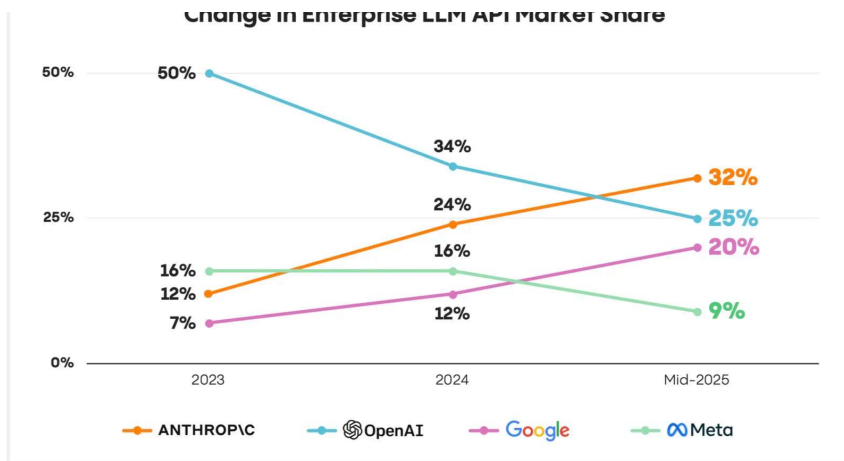
\*Non-consumer companies, including both enterprises and startups

## Anthropic Surpasses OpenAI in Enterprise Usage

By the end of 2023, [OpenAI](#) commanded **50%** of the enterprise LLM market, but its early lead has eroded. Today, it captures just **25%**<sup>3</sup> of enterprise usage—half of what it held two years ago.

Anthropic is the new top player in enterprise AI markets with **32%**, ahead of OpenAI and [Google](#) (**20%**), which has shown strong growth in recent months. [Meta](#)'s [Llama](#) holds **9%**, while [DeepSeek](#), despite its high-profile launch at the beginning of the year, accounts for just **1%**.

By using this website, you agree to our use of cookies. We use cookies to provide you with a great experience and to help our website run effectively.



© 2025 Menlo Ventures

The momentum that carried Anthropic to the top of the LLM leaderboard began in earnest with the release of Claude Sonnet 3.5 in June 2024. Momentum accelerated with Claude Sonnet 3.7 in February 2025, which introduced the first real glimpse of an agent-first LLM. By May 2025, [Claude Sonnet 4](#), [Opus 4](#), and [Claude Code](#) cemented Anthropic's lead.

Three industry-defining trends fueled Anthropic's momentum:

### 1. Code generation became AI's first killer app.

Claude quickly became the developer's top choice for code generation, capturing **42%** market share, more than double OpenAI's (**21%**). In just one year, Claude helped transform a single-product space ([GitHub Copilot](#)) into a **\$1.9 billion** ecosystem. The release of Claude Sonnet 3.5 in June 2024 demonstrated how breakthroughs at the model layer can move application markets, making possible entirely new categories like AI IDEs ([Cursor](#), [Windsurf](#)), app builders ([Lovable](#), [Bolt](#), [Replit](#)), and enterprise coding agents (Claude Code, [All Hands](#)).

### 2. Reinforcement learning with verifiers is the new path to scaling intelligence.

In 2024, the primary way to scale intelligence was by pre-training larger and

By using this website, you agree to our use of cookies. We use cookies to provide you with a great experience and to help our website run effectively.



### 3. Training models as agents to use tools makes them far more useful.

LLMs were initially designed to provide complete answers in a single response. However, enabling them to think step-by-step, reason through problems, and use external tools across multiple interactions—creating what’s known as an agent—makes them dramatically more effective for real-world applications. 2025 has become known as the “year of agents.” Anthropic led the way in training models to iteratively improve their responses and integrate tools like search, calculators, coding environments, and other resources through MCP ([model context protocol](#)), significantly boosting both their capabilities and user adoption.

## Open-Source Adoption in the Enterprise Flattens

**Thirteen percent** of AI workloads today use open-source models, down slightly from **19%** six months ago.<sup>4</sup> The market leader remains the popular Llama model by Meta, though the Llama 4 launch in April underwhelmed in real-world settings.

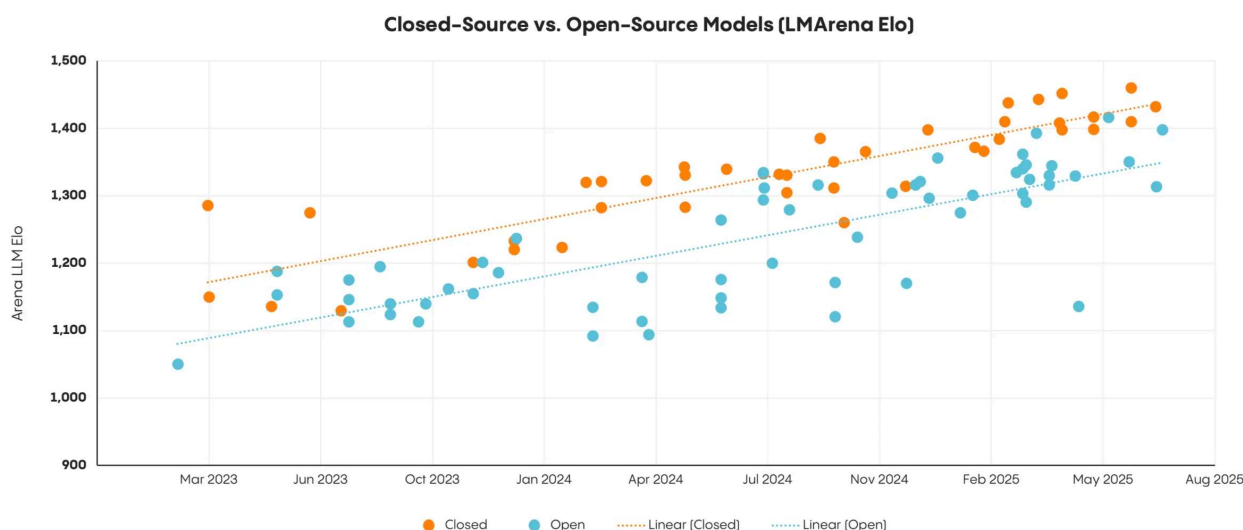
The landscape has stayed active, with notable launches from DeepSeek (V3, R1), [Bytedance Seed](#) (Doubao), [Minimax](#) (Text 1), [Alibaba](#) (Qwen 3), [Moonshot AI](#) (Kimi K2), and [Z AI](#) (GLM 4.5) in the last six months. You can try all of these in one API on [OpenRouter](#).\*.

Open-source models offer clear enterprise advantages: greater customization, potential cost savings, and the ability to deploy within private cloud or on-premises environments. But despite these benefits and recent improvements, open-source has continued to trail frontier, closed-source models in performance by nine to 12 months.

By using this website, you agree to our use of cookies. We use cookies to provide you with a great experience and to help our website run effectively.



## Closed-Source vs. Open-Source Models



© 2025 Menlo Ventures

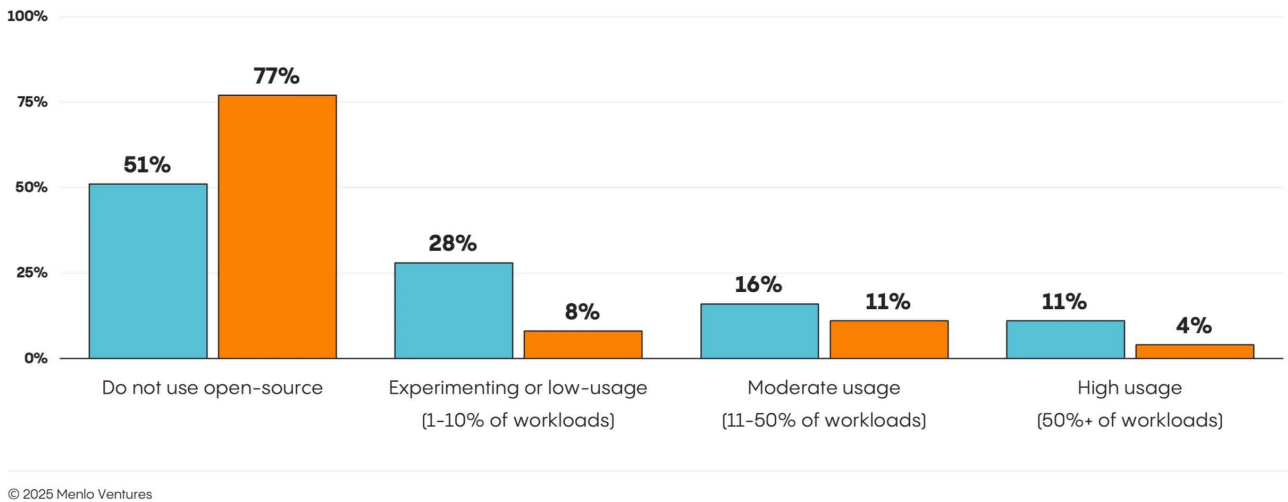
Source: LMArena

It's not just enterprises. Fewer startups adopt open-source models for these reasons, too. As one respondent put it:

Currently, 100% of our production workloads are running on closed-source models. We initially started with Llama and DeepSeek for POCs, but they couldn't keep up with the performance of closed-source over time."

By using this website, you agree to our use of cookies. We use cookies to provide you with a great experience and to help our website run effectively.





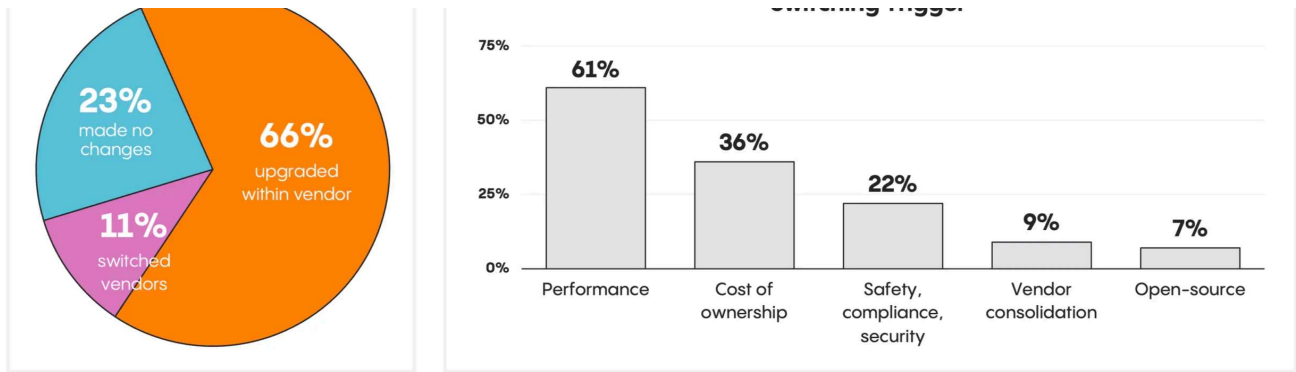
## Enterprises Switch Models for Performance, Not Price

**Switching between vendors is relatively easy, but increasingly rare.**

Most teams remain with their provider and simply upgrade to the newest model as it becomes available. Once builders commit to a platform, they tend to stay, but move quickly to upgrade to newer, higher-performing models when they're released.

According to our survey: **66%** of builders upgraded models within their existing provider, while **23%** did not switch models at all this past year. Only **11%** switched vendors.

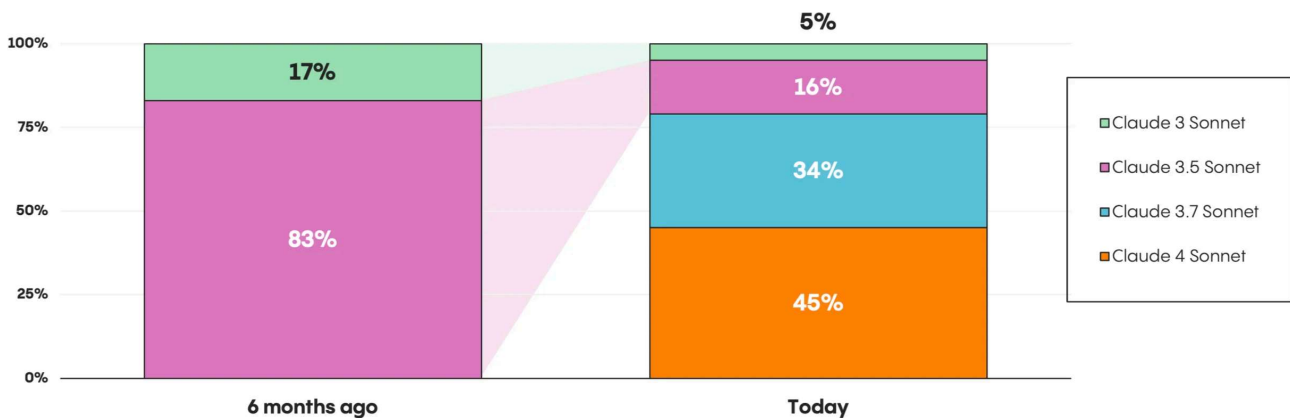
By using this website, you agree to our use of cookies. We use cookies to provide you with a great experience and to help our website run effectively.



© 2025 Menlo Ventures

**Performance drives decisions.** Builders consistently choose frontier models over cheaper, faster alternatives. They prioritize and pay for performance. When new models come out, switching happens in weeks. Within one month of Claude 4's release, for instance, Claude 4 Sonnet captured **45%** of Anthropic users while Sonnet 3.5 share decreased from **83%** to **16%**.

### Builders Choose Frontier Models

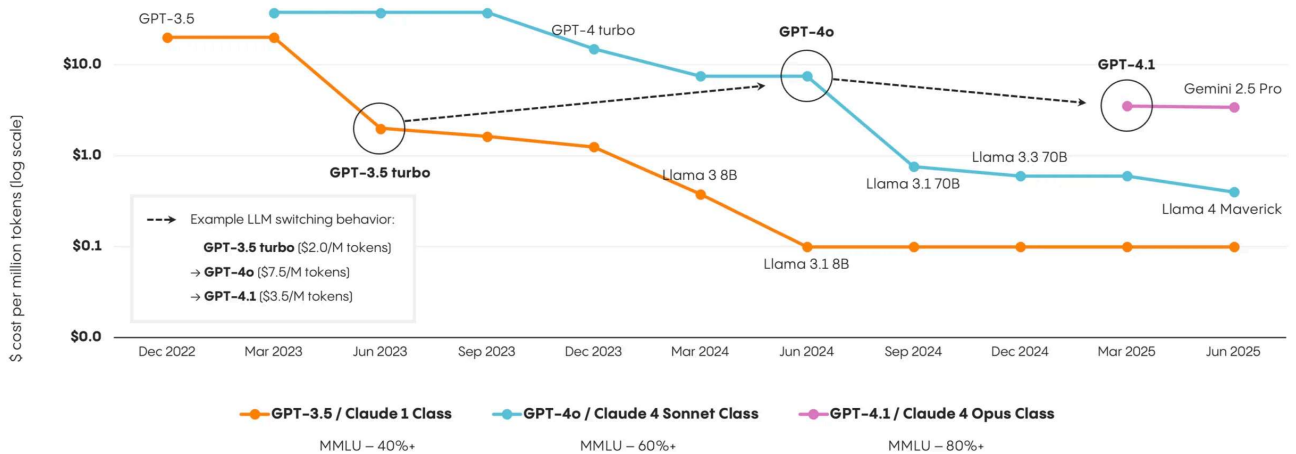


© 2025 Menlo Ventures

This creates an unexpected market dynamic: Even as individual models drop **10x** in price, builders don't capture savings by using older models; they just move en masse to the best performing one.

By using this website, you agree to our use of cookies. We use cookies to provide you with a great experience and to help our website run effectively.





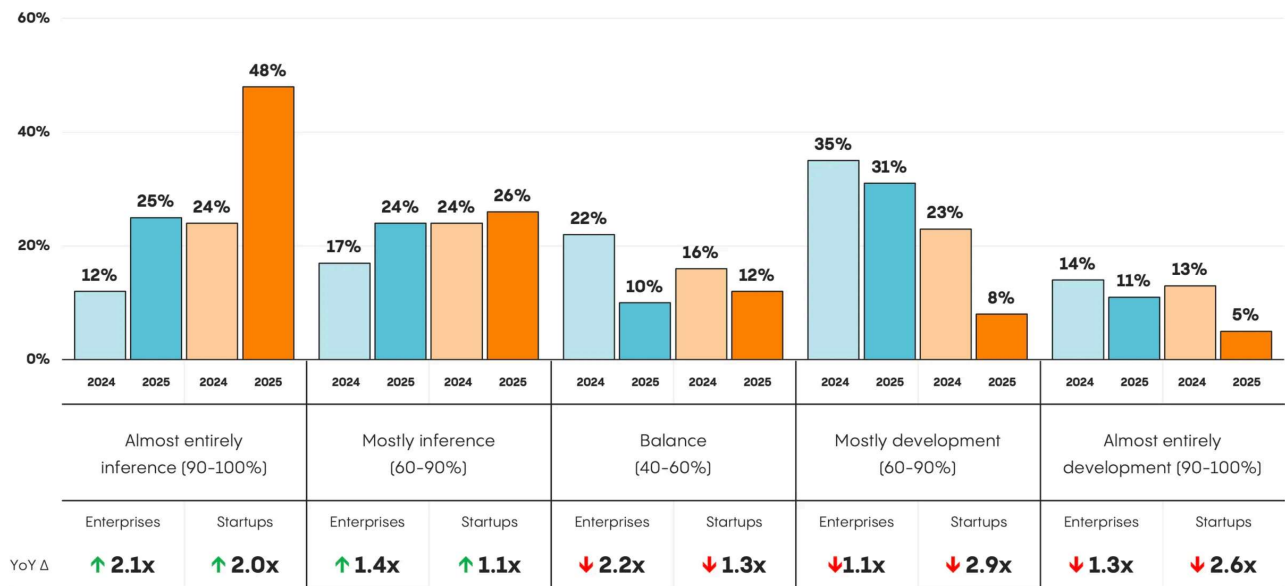
© 2025 Menlo Ventures

Source: Artificial Analysis

## AI Spend Is Moving from Training to Inference

Spending on compute is steadily shifting from building and training models to inference, with models actually running in production. This shift is most pronounced among startups: **74%** of builders now say the majority of their workloads are inference, up from **48%** a year ago. Large enterprises are not far behind. Nearly half (**49%**) report that most or nearly all of their compute is inference-driven—up from **29%** last year.

By using this website, you agree to our use of cookies. We use cookies to provide you with a great experience and to help our website run effectively.



© 2025 Menlo Ventures

## Where We Go from Here

Predicting the future of AI can be a fool's errand. The market changes by the week, with exciting new model launches, advancements in foundation model capabilities, and plunging costs. What has become clear, though, is that conditions are ripe for a new generation of enduring AI businesses to be built on top of today's foundational building blocks.

At Menlo Ventures, we've been partnering with founders building at the AI infrastructure layer for years, including [Anthropic](#), [Cleanlab](#), [Goodfire](#), [Mercor](#), [OpenRouter](#), [Pinecone](#), and [Unstructured](#). If you're creating infrastructure, tooling, and applications for the AI era, we'd love to hear from you.

### 1. Our LLM market sizing excludes frontier AI lab revenue from consumer-facing

By using this website, you agree to our use of cookies. We use cookies to provide you with a great experience and to help our website run effectively.



2. This report summarizes data from a survey of 150 technical decision-makers at enterprises and startups building AI applications, conducted from June 30 to July 10, 2025. Enterprises are defined as organizations with 5,000 or more employees. Startups included in the sample have raised at least \$5 million in venture funding. Across this foundational data, we overlaid our perspective and insights as active investors in the space. ↩
3. LLM market share reflects the proportion of production AI usage, not spend. Survey respondents reported the share of their AI workloads using each model. Responses were weighted based on each enterprise and startup application's scale. ↩
4. Source: Menlo Ventures' [2024: The State of Generative AI in the Enterprise](#), November 2024 ↩

\*Backed by Menlo Ventures

## Tim Tully

Tim is a partner at Menlo Ventures focused on early-stage investments that speak to his passion for AI/ML, the new data stack, and the next-generation cloud. His background as a technology builder, buyer, and seller informs his investments in companies like Pinecone, Neon, Edge Delta, JuliaHub, TruEra, ClearVector, Squint, and...

## Joff Redfern

A self-described "tall, slightly nerdy product guy," Joff was previously the Chief

By using this website, you agree to our use of cookies. We use cookies to provide you with a great experience and to help our website run effectively.



Deedy is a principal at Menlo Ventures focused on early-stage investments in AI/ML, next-generation infrastructure, and enterprise software. Having been an engineer and product leader at both a successful startup and large public companies, Deedy is well-equipped to help technical founders navigate how to build and scale enduring tech companies....

---

## Derek Xiao

As an investor at Menlo Ventures, Derek concentrates on identifying investment opportunities across the firm's thesis areas, covering AI, cloud infrastructure, digital health, and enterprise SaaS. Derek is especially focused on investing in soon-to-be breakout companies at the inflection stage. Derek joined Menlo from Bain & Company where he worked...

# Related Perspectives

PORTFOLIO PERSPECTIVE

## 2025: The State of Consumer AI



By using this website, you agree to our use of cookies. We use cookies to provide you with a great experience and to help our website run effectively.



---

PORTFOLIO   FUNDING

## **Investing in OpenRouter, the One API for All AI**

By using this website, you agree to our use of cookies. We use cookies to provide you with a great experience and to help our website run effectively.



---

PORTFOLIO   FUNDING

## Leading Goodfire's \$50M Series A to Interpret How AI Models Think

By using this website, you agree to our use of cookies. We use cookies to provide you with a great experience and to help our website run effectively.





---

PORTFOLIO PERSPECTIVE

## **Security for Agents and Agents for Security: The Next Cybersecurity Frontier**

By using this website, you agree to our use of cookies. We use cookies to provide you with a great experience and to help our website run effectively.



---

PORTFOLIO   FUNDING

## Why We're Betting Wispr Will Kill the Keyboard

By using this website, you agree to our use of cookies. We use cookies to provide you with a great experience and to help our website run effectively.



VIEW ALL

Company

ABOUT US

TEAM

PORTFOLIO

CONTACT

PERSPECTIVE

Menlo Programs

ANTHOLOGY FUND

MENLO LABS

MENLO FELLOWS

Focus Area

AI

BIO + HEALTHCARE

CONSUMER

CYBERSECURITY

FINTECH

INFRASTRUCTURE

ROBOTICS + HARDWARE

SAAS

SUPPLY CHAIN + AUTOMATION

By using this website, you agree to our use of cookies. We use cookies to provide you with a great experience and to help our website run effectively.



CAREERS

[TERMS OF USE AND CONDITIONS](#)

[PRIVACY POLICY](#)

[SITE BY TAKEOFF](#)

© 2025 Menlo Ventures. All rights reserved.

By using this website, you agree to our use of cookies. We use cookies to provide you with a great experience and to help our website run effectively.