MENLO VENTURES

2025: The State of Generative AI in the Enterprise

December 9, 2025

Tim Tully, Joff Redfern, Deedy Das, and Derek Xiao

AI Boom vs. Bubble

For all the fears of over-investment, AI is spreading across enterprises at a pace with no precedent in modern software history.

For nearly three years, AI enjoyed unwavering confidence and record capital flows. The surge crowned Nvidia the world's most valuable company.[1] Foundation models announced close to **$1 trillion** in AI infrastructure commitments. Venture funding surged back toward all-time highs, with nearly half of it concentrated in just a handful of frontier AI labs.

Enterprise AI Is the Fastest-Scaling Software Category in History

Now capturing **6%** of the ~$300B global SaaS market*

*Source: Gartner

Enterprise AI has surged from $1.7B to $37B since 2023, now capturing 6% of the global SaaS market and growing faster than any software category in history.

1. Reuters, "Nvidia Breaches $5 Trillion Market Cap," October 29, 2025, https://www.reuters.com/business/view-nvidia-breaches-5-trillion-market-cap-2025-10-29/; Companies Market Cap, "NVIDIA Market Cap," accessed December 2025, https://companiesmarketcap.com/nvidia/marketcap

Then the euphoria peaked. An MIT study[2] claiming that **95%** of generative AI initiatives fail rattled markets over the summer, exposing how quickly sentiment could shift beneath the weight of AI's massive capex spend. The whispers of a bubble became a din.

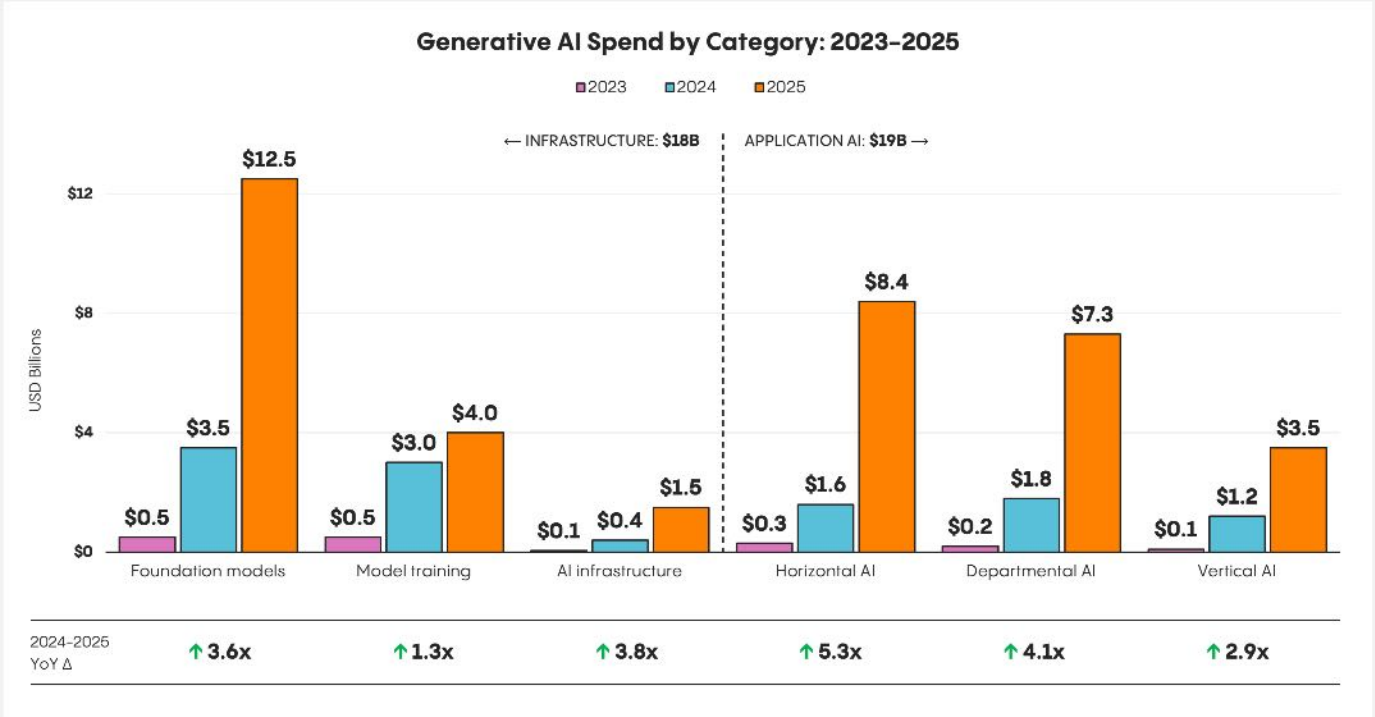
The concerns aren't unfounded given the magnitude of the numbers being thrown around. But the demand side tells a different story: Our latest market data shows broad adoption, real revenue, and productivity gains at scale, signaling a boom versus a bubble.

For Menlo's third annual State of Generative AI in the Enterprise report, we surveyed ~500 U.S. enterprise decision-makers and combined their insights with a bottoms-up model of the generative AI market spanning model APIs, infrastructure, and applications. Since 2023, our team has tracked AI's evolution from early experiments to broad enterprise deployment, giving us a multi-year view of the speed and scale of this transformation.

Follow the Money: Where Do Enterprise Dollars Flow?

Our data indicates companies spent **$37 billion** on generative AI in 2025,[3] up from **$11.5 billion**[4] in 2024, a **3.2x** year-over-year increase. The largest share, **$19 billion**, went to the user-facing products and software that leverage underlying AI models, aka the application layer. This represents more than **6%** of the entire

Where Does the GenAI Budget Go?



In 2025, more than half of enterprise AI spend went to AI applications, indicating that modern enterprises are prioritizing immediate productivity gains vs. long-term infrastructure bets.

2. MIT MLQ AI, "State of AI in Business 2025 Report," 2025, https://mlq.ai/media/quarterly_decks/v0.1_State_of_AI_in_Business_2025_Report.pdf
3. Generative AI spending includes dollars that went to foundation models, model training, AI infrastructure, and AI applications from both startups and incumbents. Note that this market sizing does not include revenue for chips (e.g., Nvidia), inference and model serving (e.g., AWS, GCP, Azure, Fireworks), or AI features built into existing software solutions (e.g., Intuit Assist). For detailed methodology, see Methodology section.
4. The estimate of $1.7 billion of generative AI spend in 2023 and $11.5 billion in 2024 excludes inference, which was previously included in Menlo Ventures' 2023 and 2024 State of Generative AI in the Enterprise reports.

software market, all achieved within three years of ChatGPT's launch.

Growth extends far beyond a handful of AI chat apps, touching every domain in the economy. By our count, there are now at least **10 products** generating over **$1 billion in ARR** and **50 products** generating over **$100 million in ARR**, led by the model APIs powering applications (Anthropic*, OpenAI, Google), but increasingly distributed across departmental solutions in coding, sales, customer support, HR, and verticals from healthcare and legal to the creator economy.

How AI Enters the Enterprise: The Path to Production

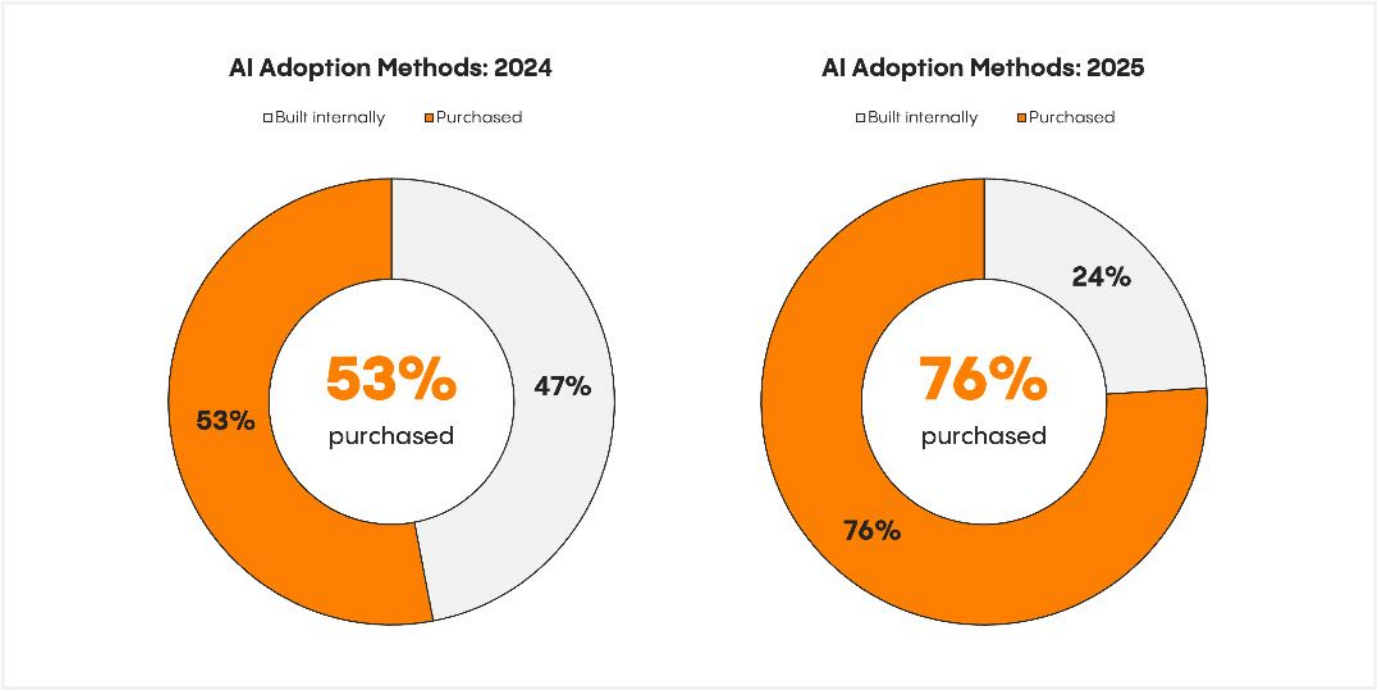
After three years, enterprise AI's path to production has taken shape. Early adopters had no playbook. Now, distinct patterns have emerged that break from traditional SaaS. Enterprises prefer buying over building, showing stronger purchase intent, and adopting AI through product-led growth at a scale rarely seen in enterprise software.

Enterprises Are Buying More Than Building

For a while, the prevailing wisdom was that enterprises would build most AI solutions themselves. Bloomberg trained BloombergGPT for finance in 2022, Walmart built Wallaby for retail in 2024. Teams were confident that, with the right data, domain expertise, and scaffolding, they could handle everything in-house.

In 2024, that confidence still showed in the data: **47%** of AI solutions were built internally, **53%** purchased.[5] Today, **76%** of AI use cases are purchased rather than built internally. Despite continued strong investments in internal builds, ready-made AI solutions are reaching

Building vs. Buying Enterprise AI Solutions



Last year, enterprises were split on building vs. buying. Today, enterprises have more ready-made AI solutions in production as their internal builds mature.

5. Menlo Ventures, "2024: The State of Generative AI in the Enterprise," November 20, 2024, https://menlovc.com/2024-the-state-of-generative-ai-in-the-enterprise/

AI Buyers Convert at Nearly 2x the Rate of Typical Software Procurement

From pilot to production: AI vs. traditional software

AI buyers convert at 47% vs. SaaS' conversion rate of 25%, indicating that AI delivers enough immediate value to short-circuit standard procurement processes.

production more quickly and demonstrating immediate value while enterprise tech stacks continue to mature.

AI Buyers Convert at Higher Rates

Enterprise buyers approach AI with notably high intent. We found that, once an organization commits to exploring an AI solution, deals convert at nearly twice the rate of traditional software: **47%** of AI deals go to production, compared to **25%** for traditional SaaS. That elevated conversion reflects strong buyer commitment and clear immediate value. Our survey data reveals that most organizations surface a long list of potential AI use cases—often **10** or more—but focus their adoption on near-term productivity gains or cost savings. While enterprises identify slightly more internal-facing use cases (**59%**) than customer-facing ones (**41%**), both categories move through the pipeline at nearly identical rates, suggesting that operational AI investments deliver value just as reliably as customer-facing innovations.

PLG: Individual Users Now Drive AI Adoption at 4x the Rate of Software

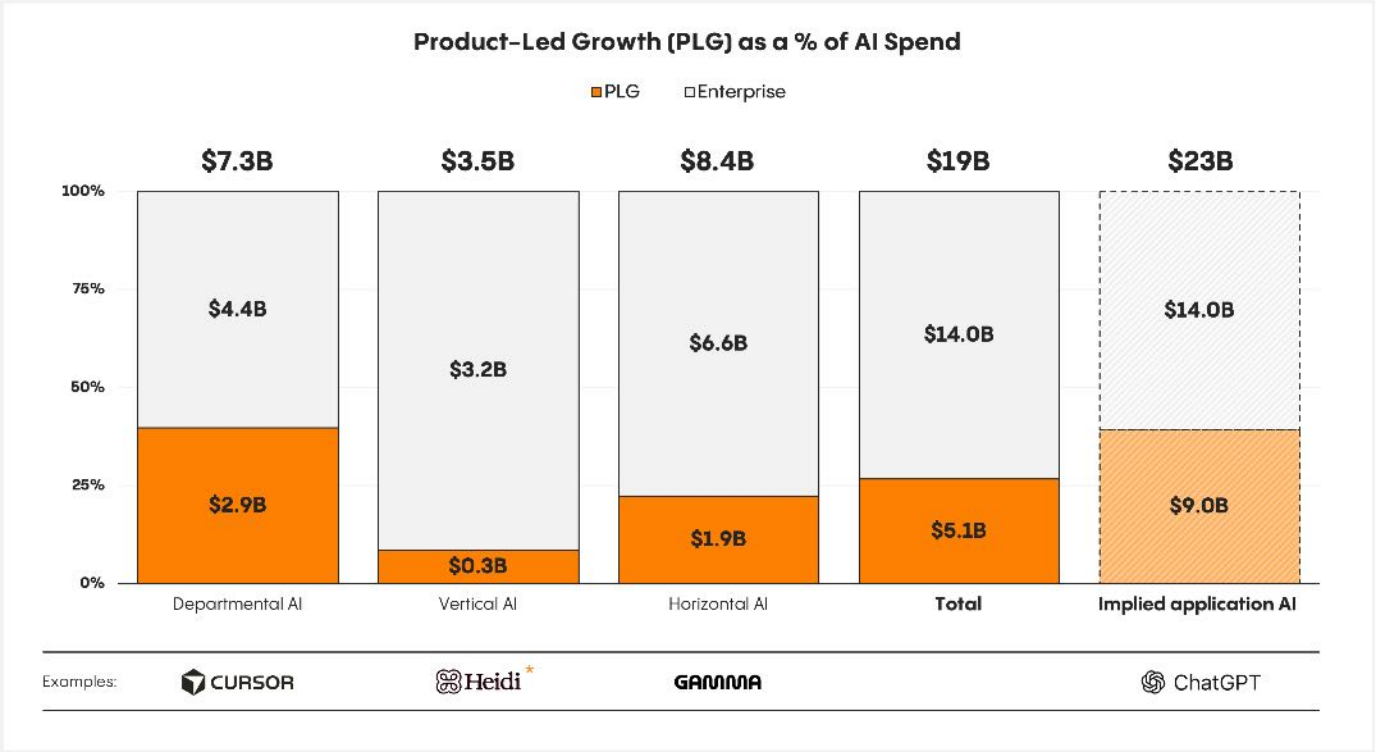
Outside of centralized procurement channels, AI solutions are increasingly finding their initial "land" in the enterprise through individual users rather than the enterprise executive. We found that **27%** of all AI application spend comes through product-led growth (PLG) motions, nearly **4x** the rate in traditional software (**7%**).

And that number is conservative. When we account for "shadow AI adoption"—employees using personal credit cards for tools like ChatGPT Plus, where roughly **27%** of usage is work-related[6]—PLG-driven tools may represent close to **40%** of application AI spend.

In AI, PLG motions are reaching enterprise scale faster and going further than in traditional SaaS. Cursor reached **$200 million** in revenue before hiring a single enterprise sales rep. n8n built its business on open-

6. Aaron Chatterji, Tom Cunningham, David J. Deming, Zoë Hitzig, Christopher Ong, Carl Shan & Kevin Wadman, How People Use ChatGPT, NBER Working Paper No. 34255, September 2025, https://doi.org/10.3386/w34255

Bottoms-Up Adoption Is a Major Driver



* Backed by Menlo Ventures

In AI, PLG motions reach and convert enterprise users far faster than traditional SaaS. Real usage proves value well before any formal contracting process begins.

source community adoption, formalizing contracts only after hundreds of employees were already active users. ElevenLabs, Gamma, and Wispr Flow* scaled in the same way.

Developers and technical teams are especially receptive to this motion. Many discover tools for individual use, prove their value in day-to-day work, and create bottom-up demand that eventually converts to enterprise contracts. Lovable, OpenRouter*, and fal follow this pattern, turning informal adoption by product managers and engineers into enterprise agreements once the tools are embedded in development workflows.

Startups vs. Incumbents: New Entrants Gain Ground in AI Apps

At the AI application layer, startups have pulled decisively ahead. This year, according to our data, they captured nearly $2 in revenue for every $1 earned by incumbents—**63%** of the market, up from **36%** last year[7] when enterprises still held the lead.

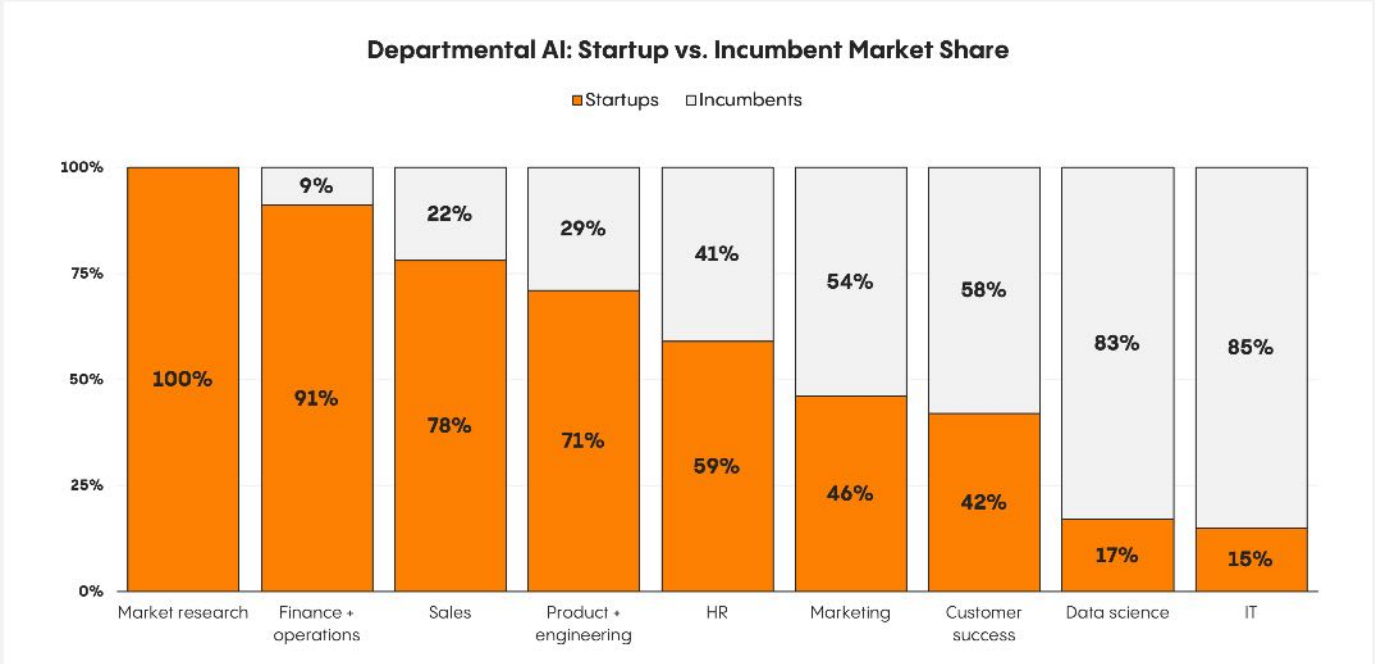
On paper, this shouldn't be happening. Incumbents have entrenched distribution, data moats, deep enterprise relationships, scaled sales teams, and massive balance sheets. Yet, in practice, AI-native startups are out-executing much larger competitors across some of the fastest-growing app categories.

7. Menlo Ventures, "2024: The State of Generative AI in the Enterprise," November 20, 2024, https://menlovc.com/2024-the-state-of-generative-ai-in-the-enterprise/

- **Product + engineering** (**71%** startup share): Code generation is the canonical example for why startups win. GitHub Copilot was the first mover and had every structural advantage, yet Cursor captured significant share by shipping better features, faster—beating Copilot to repo-level context, multi-file editing, diff approvals, and natural language commands. Cursor's model-agnostic approach let developers adopt frontier models like Claude Sonnet 3.5 the moment they launched, rather than being limited by Microsoft's partner choices. That product velocity created a PLG flywheel: Cursor won the ground game with individual developers, who then brought it into the enterprise.
- **Sales** (**78%** startup share): AI-native startups like Clay and Actively win by attacking workflows Salesforce doesn't own: research, personalization, and enrichment, which rely heavily on unstructured signals (web, social, email) that sit outside the CRM. By owning these off-CRM surfaces and expanding downstream, they become the AI layer reps actually interface with—disintermediating the legacy system of record in the near term and positioning themselves to potentially become the system of record in the long term.
- **Finance + operations** (**91%** startup share): In highly regulated domains like finance, incumbents like Intuit QuickBooks face high demands for accuracy that slow their ability to ship AI-native workflows. Although the total dollars here are still small, this paralysis creates a vacuum for startups like Rillet, Campfire, and Numeric* downmarket to build AI-first ERPs with real-time automation and intelligent workflows—winning because the incumbent cannot move fast enough to deliver a credible next-gen product.

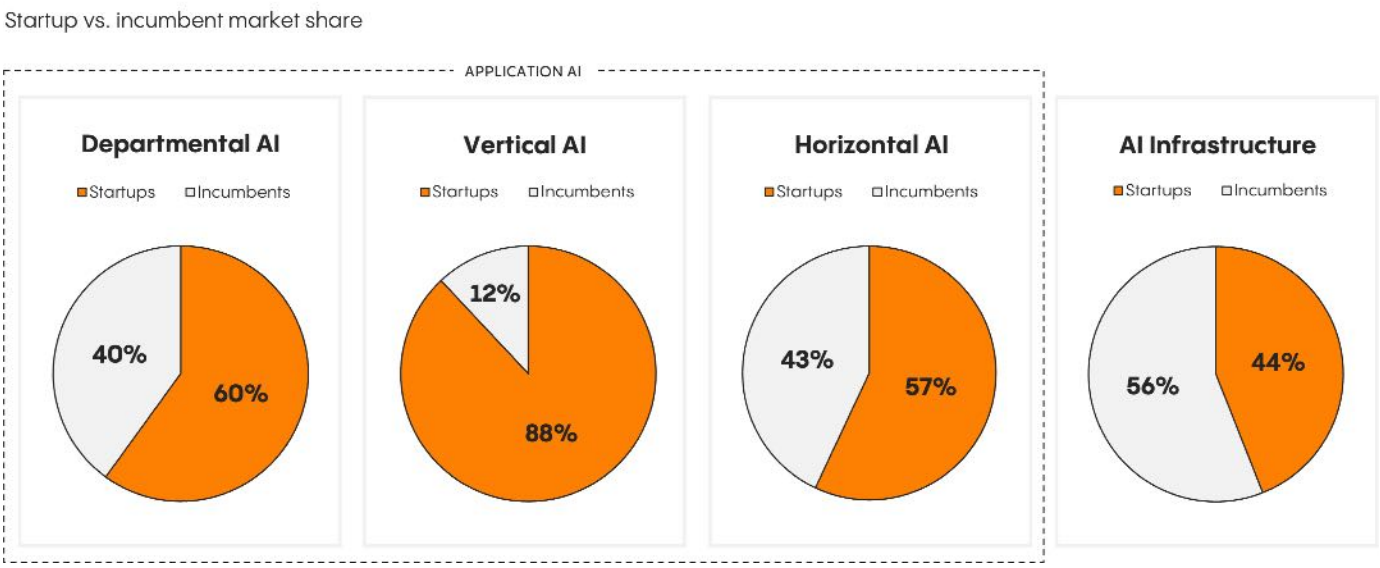
The chart below shows how this dynamic varies across enterprise departments, each with its own function-specific tooling. Teams grappling with fragmented, data-heavy workflows that lend themselves to automation

Startup vs. Incumbent Market Share by Department



AI startups thrive in agile departments like market research, sales, marketing, and product. Incumbents hold their ground in IT and data science, where reliability and deep integrations outweigh speed.

Startups Are Winning Across Application AI, Trail in Infrastructure



Startups dominate AI applications, earning nearly $2 for every $1 incumbents earn, while enterprise infrastructure spend continues to favor incumbents.

lead AI adoption. Incumbents remain stronger where reliability, integration depth, and existing system dependencies outweigh the benefits of rapid iteration.

The story changes as we move down the stack. At the infrastructure layer, the picture is more mixed. According to our data, incumbents hold **56%** of the market as many AI app builders continue building on the data platforms they've trusted for years. Although new AI-native infrastructure companies like Temporal, Supabase, Neon*, and Pinecone* are seeing impressive growth, incumbents like Databricks, Snowflake, MongoDB, and Datadog have enjoyed just as meaningful re-acceleration—as even new AI-native app builders are still primarily choosing existing platforms to manage their data, orchestrate workflows, and monitor operations.

AI Applications: A $19 Billion Market

The application layer captured **$19 billion** in 2025, more than half of all generative AI spending. This spend segments into three categories:

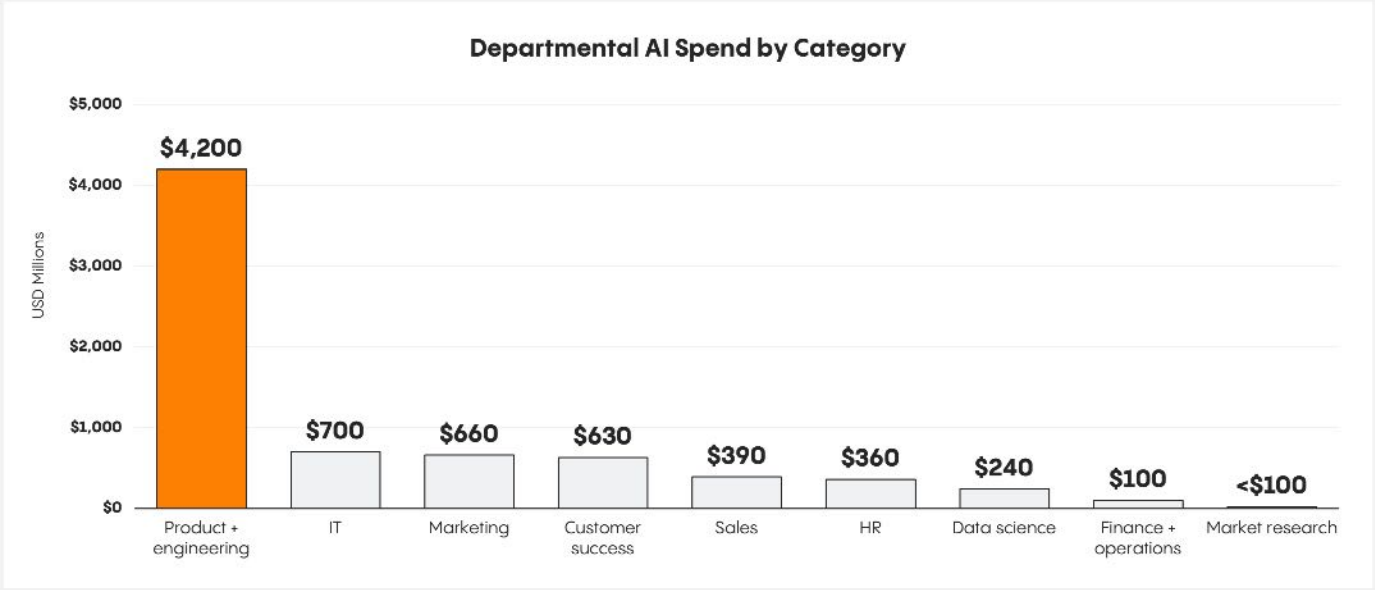
- **Departmental AI** (**$7.3 billion**), built for specific job roles like software development or sales;
- **Vertical AI** (**$3.5 billion**), targeting specific industries like healthcare or finance; and
- **Horizontal AI** (**$8.4 billion**), increasing productivity across all functions.

Departmental AI: Coding Is Generative AI's First "Killer Use Case"

Departmental AI spending hit **$7.3 billion** in 2025, up **4.1x** year over year. Coding is the clear standout at **$4.0 billion** (**55%** of departmental AI spend), making it the largest category across the entire application layer; the rest spans IT (**10%**), marketing (**9%**), customer success (**9%**), design (**7%**), and HR (**5%**).

Code became AI's first true "killer use case" as models reached economically meaningful performance—with Anthropic's Sonnet 3.5 triggering the category's initial breakout in mid-2024. Adoption followed soon after; **50%** of developers now use AI coding tools daily (**65%**

Coding Dominates $7.3B Departmental AI Market

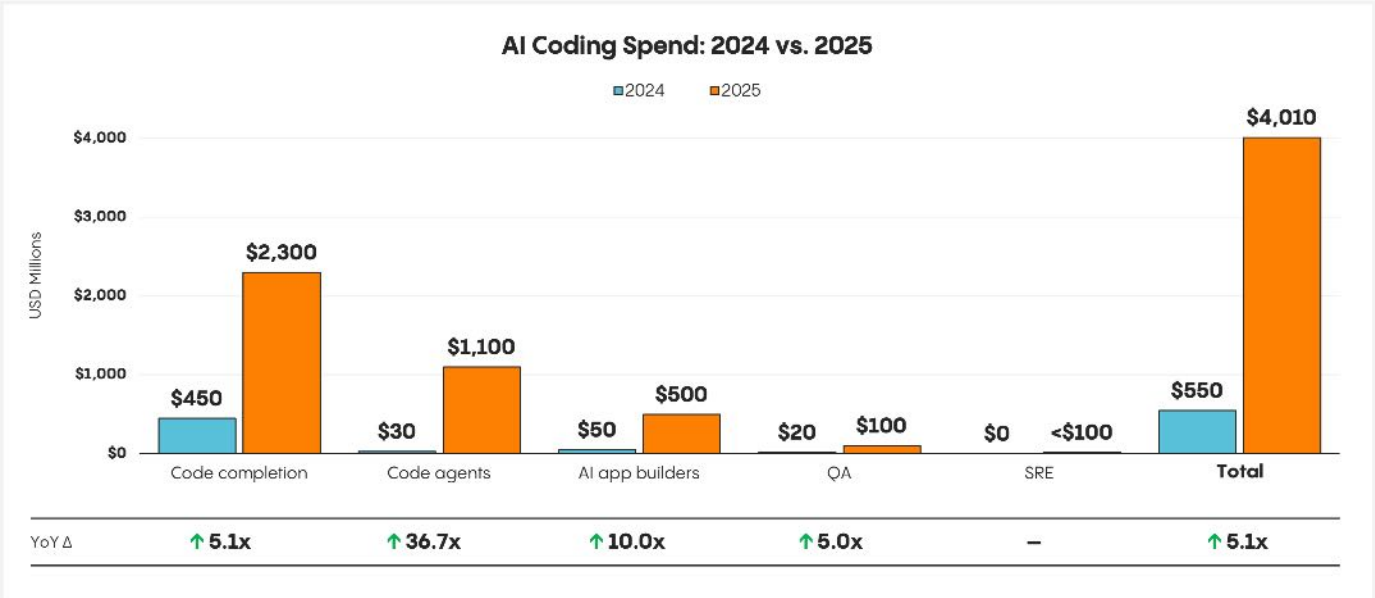


Coding has become the breakout use case in departmental AI. Investment is concentrated where the impact is most immediate: product and engineering teams now account for the vast majority of spend.

in top-quartile orgs). Code completion grew to **$2.3 billion**, while code agents and AI app builders exploded from near-zero. Teams report **15%+** velocity gains as they've adopted AI tools across the software development lifecycle: from

AI's First "Killer Use Case": Coding Is a $4B Market

The market grew 5x as coding capabilities evolved from simple autocomplete to autonomous development

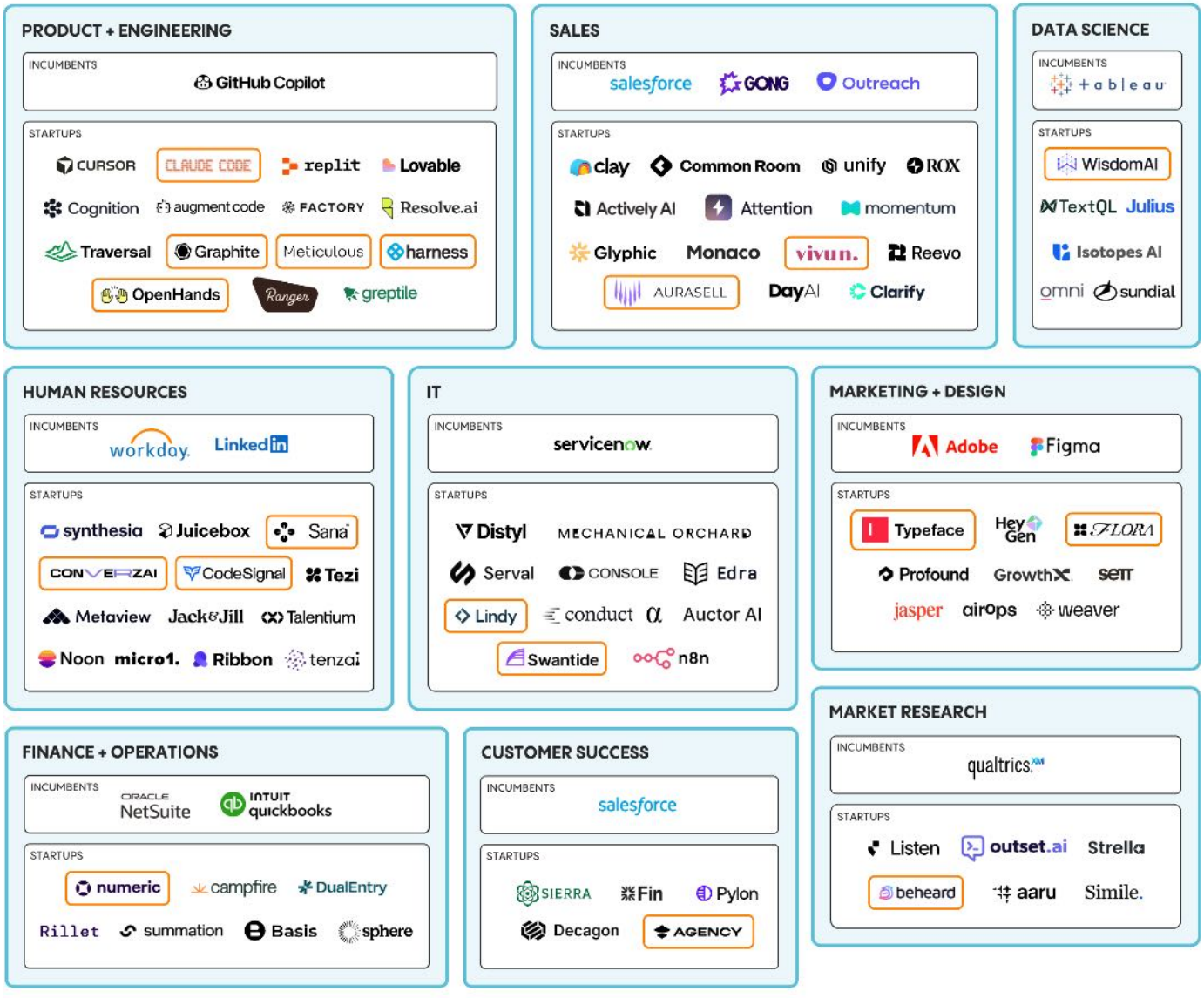


The sharp jump from $550M to $4B in 2025 reflects a shift in capability: Models can now interpret entire codebases and execute multi-step tasks. Coding moves from a point solution to an end-to-end automation category.

prototyping (Lovable) to code refactoring (Open Hands*), design-to-code (Weaver), QA (Meticulous*), PRs (Graphite*), site reliability engineering (Resolve), and deployment (Harness*).

Although coding captures more than half of departmental AI spend at **$4 billion**, the technology is gaining traction across many enterprise departments. IT operations tools reached **$700 million** as teams automated incident response and infrastructure management. Marketing platforms hit **$660 million**, driven by content generation and campaign optimization. Customer success tools captured **$630 million**, with AI handling ticket routing, sentiment analysis, and proactive outreach. Each of these categories targets repetitive workflows where productivity gains are immediate and measurable. The market map below shows which players have emerged across functions to capture part of the **$7.3 billion** enterprise investment in departmental AI.

Departmental AI Market Map



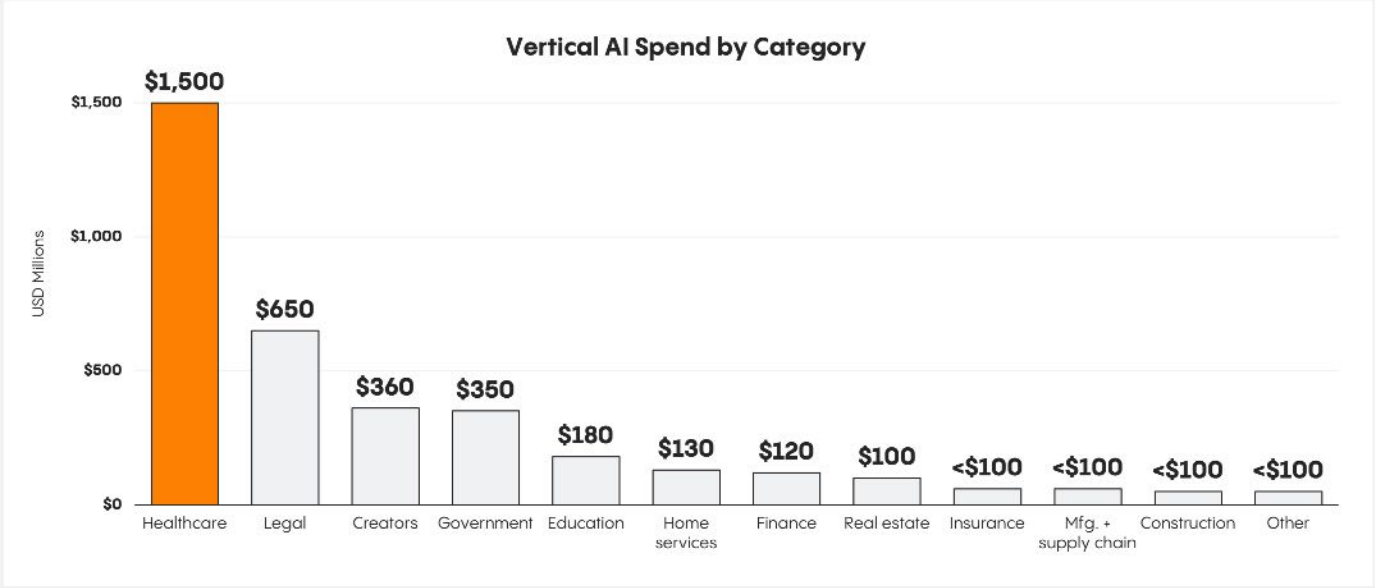
AI-native startups are rapidly emerging across every job function, capturing a meaningful share of the $7.3B spent on departmental AI in 2025.

Backed by Menlo Ventures

Vertical AI: Healthcare Leads Adoption

Vertical AI solutions captured **$3.5 billion** in 2025, nearly **3x** the **$1.2 billion** invested in 2024. When segmented by industry, healthcare alone captures nearly half of all vertical AI spend—approximately **$1.5 billion**, more than tripling from **$450 million** the year prior and exceeding the next four verticals combined.

Healthcare Dominates $3.5B Vertical AI Market



$3.5B was invested across vertical sectors this year, nearly 3x last year's spend. Healthcare represents $1.5B of that total, capturing 43% of the market and outspending the next four verticals combined.

Healthcare moves slowly, bogged down by long procurement cycles and regulatory headwinds. But after years of rising administrative burden, shrinking margins, and chronic staffing shortages, health systems became one of the strongest sources of demand for AI automation anywhere in the economy.

The bulk of spend concentrates in administrative and clinical-adjacent workflows, led by ambient scribes. The scribe market reached **$600 million** in 2025 (**+2.4x** YoY),[8] minting two new unicorns (Abridge and Ambience) alongside the market leader, Nuance's DAX Copilot. Because clinicians spend roughly one hour documenting for every five hours of care, scribes that reduce documentation time by more than **50%** can dramatically reduce administrative burden and free doctors to practice at the top of their license.

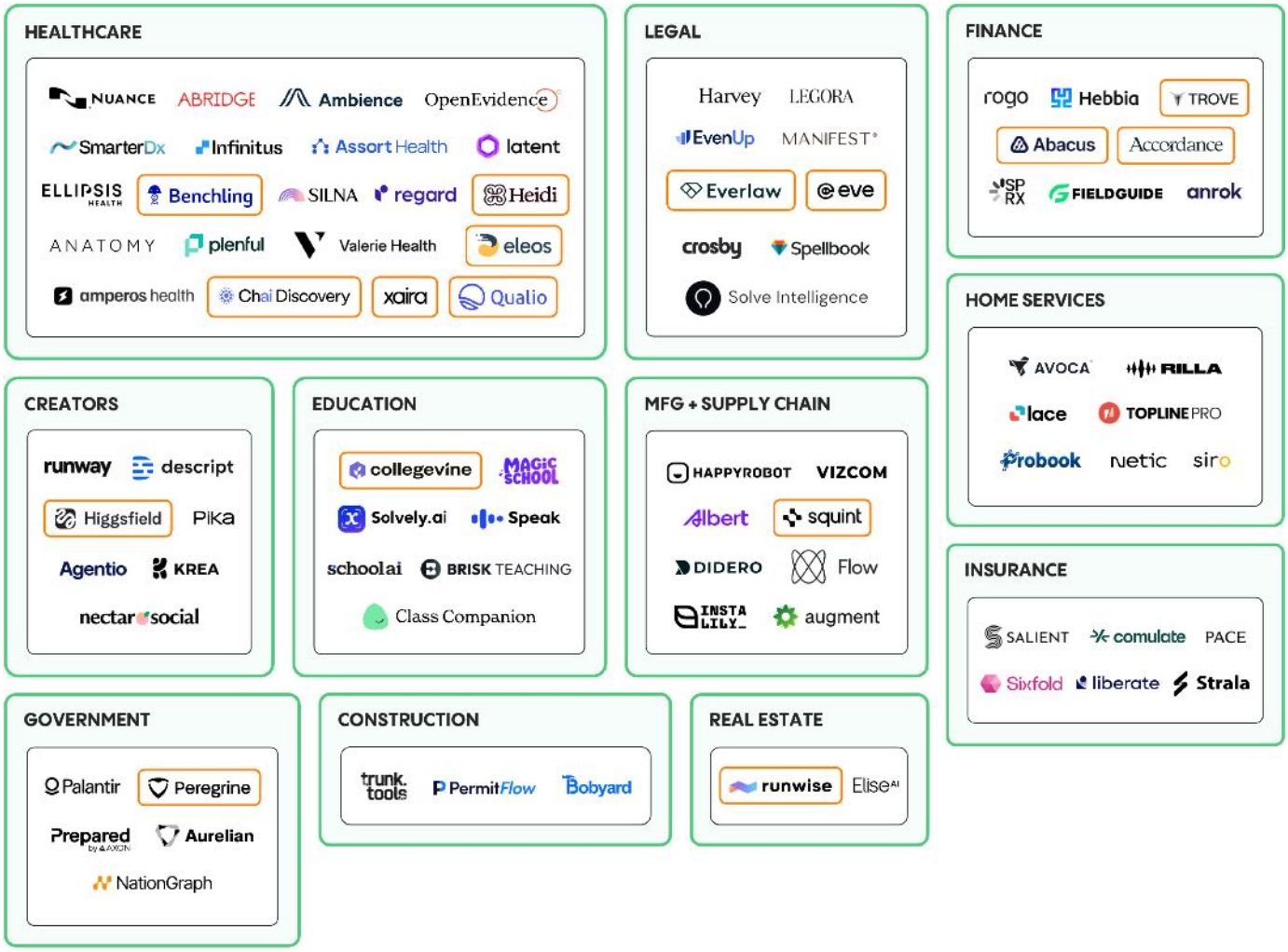
For deeper analysis of how AI is transforming healthcare, see our 2025: The State of AI in Healthcare report on adoption trends, budget shifts, and the areas where health systems are already seeing meaningful ROI.

Beyond healthcare, AI is beginning to take hold across nearly every sector of the economy. Led by companies like Eve*, legal has grown into a **$650 million** market; creator tools into **$360 million**; and government into **$350 million**. Adoption is strongest in industries historically underserved by software: fields defined by manual,

8. Menlo Ventures, "2025: The State of AI in Healthcare," October 21, 2025, https://menlovc.com/perspective/2025-the-state-of-ai-in-healthcare/

unstructured workflows that once depended on human services but can now be automated with generative AI. The market map below highlights the companies building across these sectors and vying for a share of the **$3.5 billion** enterprises poured into vertical AI this year.

Vertical AI Market Map

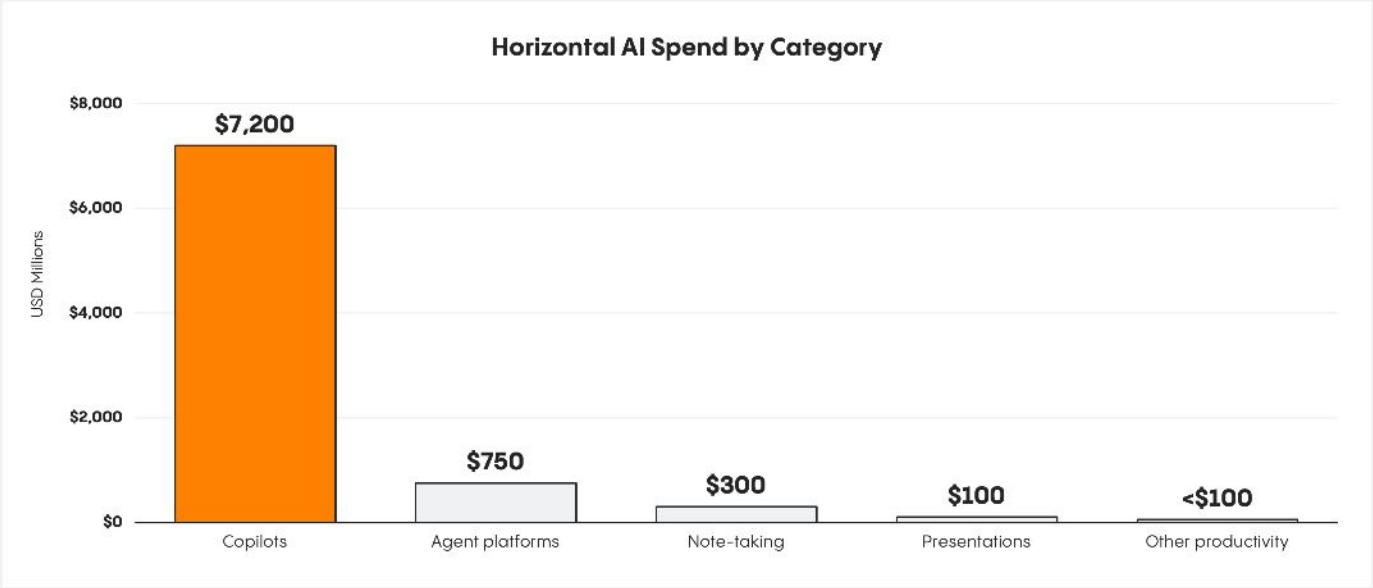


Vertical AI has become a $3.5B category in 2025, triple the dollars invested last year. These companies illustrate how AI-native software is emerging to serve every sector of the economy.

Horizontal AI: Copilots Dwarf Agent Spend

At **$8.4 billion**, horizontal AI remains the largest and fastest-growing category in the application layer, expanding **5.3x** year over year. Copilots dominate with **86%** share (**$7.2 billion**)—led by ChatGPT Enterprise, Claude for Work, and Microsoft Copilot. Agent platforms such as Salesforce Agentforce, Writer, and Glean capture another **10%** (**$750 million**), while personal productivity tools like Granola and Fyxer account for the remaining **5%** (**$450 million**).

Copilots Are 10x Bigger Than Agents—For Now



General-purpose copilots dominate today, but as agents become more powerful, we can expect a shift from assistance to automation.

AI Infrastructure: $18 Billion for "Picks and Shovels"

Our data shows the infrastructure layer captured **$18 billion** in 2025—the other half of all generative AI spending and up **2.0x** from **$9.2 billion** in 2024. This spend segments into three categories:

- Foundation model APIs (**$12.5 billion**) power the intelligence behind all AI applications.
- Model training infrastructure (**$4.0 billion**) enables frontier labs and enterprises to train and adapt models.
- AI infrastructure (**$1.5 billion**) manages the storage, retrieval, and orchestration of data that connects LLMs to enterprise systems.

LLM Market Share: Anthropic Extends Its Lead in the Enterprise

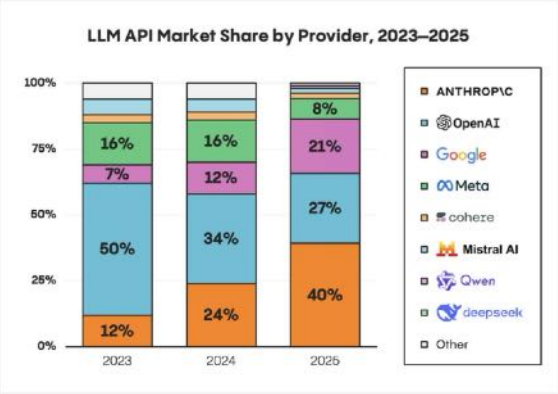
The foundation model landscape shifted decisively this year when Anthropic surprised industry watchers by unseating OpenAI as the enterprise leader. We estimate Anthropic now earns **40%** of enterprise LLM spend,[9] up from **24%** last year and **12%** in 2023. Over the same period, OpenAI lost nearly half of its enterprise share, falling to **27%** from **50%** in 2023. Google also saw significant gains, increasing its enterprise share from **7%** in 2023 to **21%** in 2025. Together, these three companies account for **88%** of enterprise LLM API usage, with the remaining **12%** spread across Meta's Llama, Cohere, Mistral, and a long tail of smaller providers.

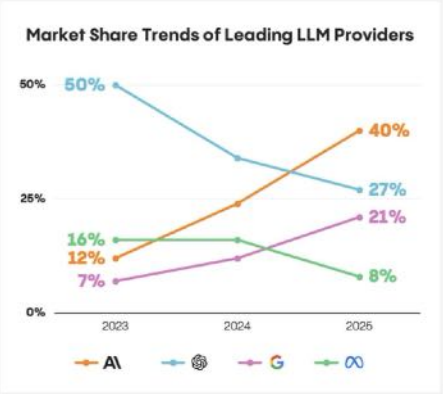
Anthropic's ascent has been driven by its remarkably durable dominance in the coding market, where it now commands an estimated **54%** market share, compared to **21%** for OpenAI. This is up from **42%** just six months ago, driven in large part by the popularity of Claude Code.

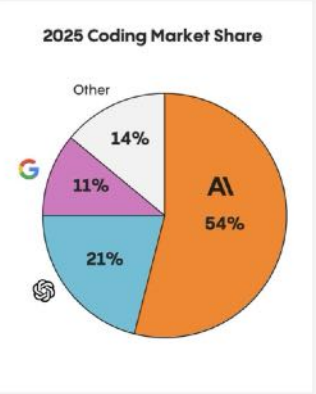
In fact, Anthropic has now had an almost unparalleled **18 months** atop the LLM leaderboards for coding, starting with the release of Claude Sonnet 3.5 in June 2024. When Google released Gemini 3 Pro in mid-

9. LLM market shares approximate dollars spent based on proportion of production API usage. Survey respondents reported the share of their AI workloads using each model. Responses were then weighted based on each enterprise and startup application's scale, and results triangulated with publicly reported financials.

Enterprise LLM API Market Share by Usage



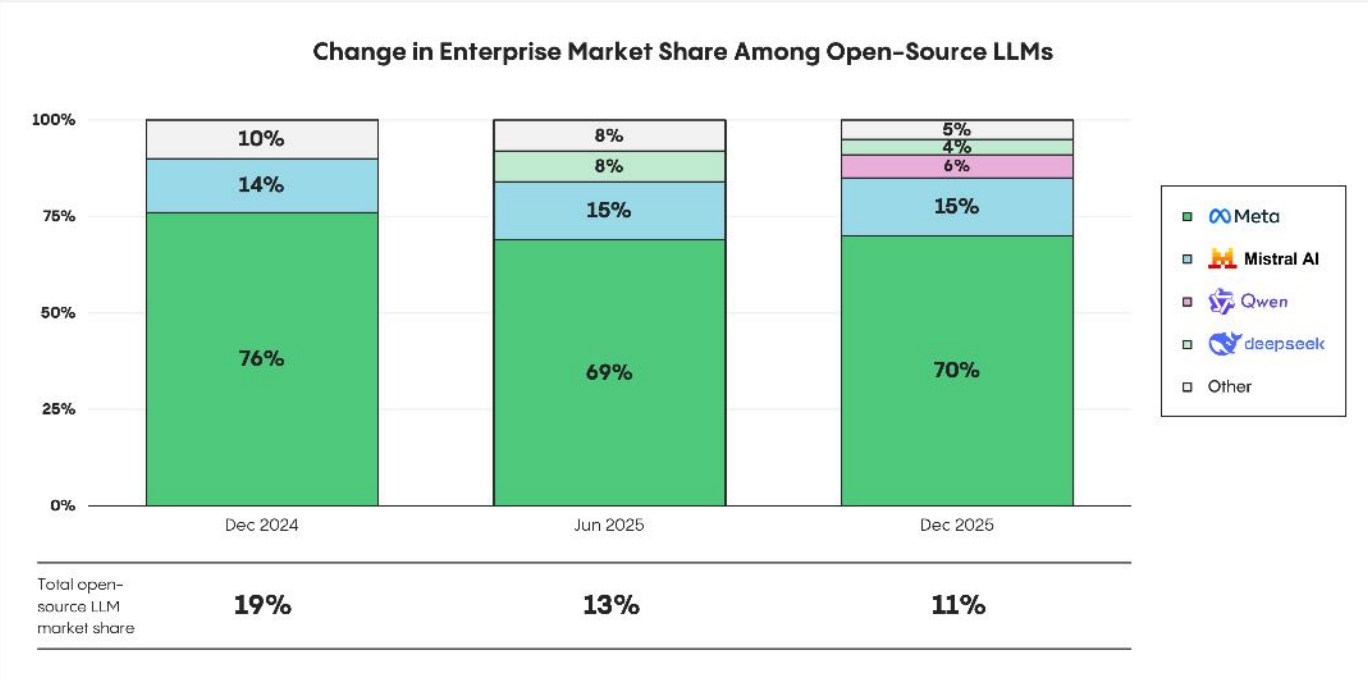




Together, these views capture how the LLM ecosystem has shifted. The stacked bars show the magnitude of shifts in market share, the trend lines highlight the momentum behind the leading providers, and the coding share underscores where competitive advantage is being won.

November 2025, its own model card[10] showed it leading most major evaluations—except SWE-bench Verified,[11] where it still trailed Claude Sonnet 4.5. Just a week later, Anthropic widened the gap again with Claude Opus 4.5,[12] which reset the high-water mark for code generation and reaffirmed Anthropic's position as the category's top performer.

Open-Source LLMs Lag in the Enterprise



Enterprises remain cautious, preferring closed-source models. Open-source LLMs hold only 11% of today's market, but developers are pushing boundaries, running Chinese models in production and testing new architectures at scale.

10. Google DeepMind, "Gemini 3 Technical Report," November 2025, https://storage.googleapis.com/deepmind-media/Model-Cards/Gemini-3-Pro-Model-Card.pdf
11. SWE-bench, "SWE-bench Verified Leaderboard," accessed December 2025, https://www.swebench.com/
12. Anthropic, "Claude 4.5," December 2025, https://www.anthropic.com/news/claude-4-5

Open-Source Models: Enterprise Adoption Lags the Broader Ecosystem

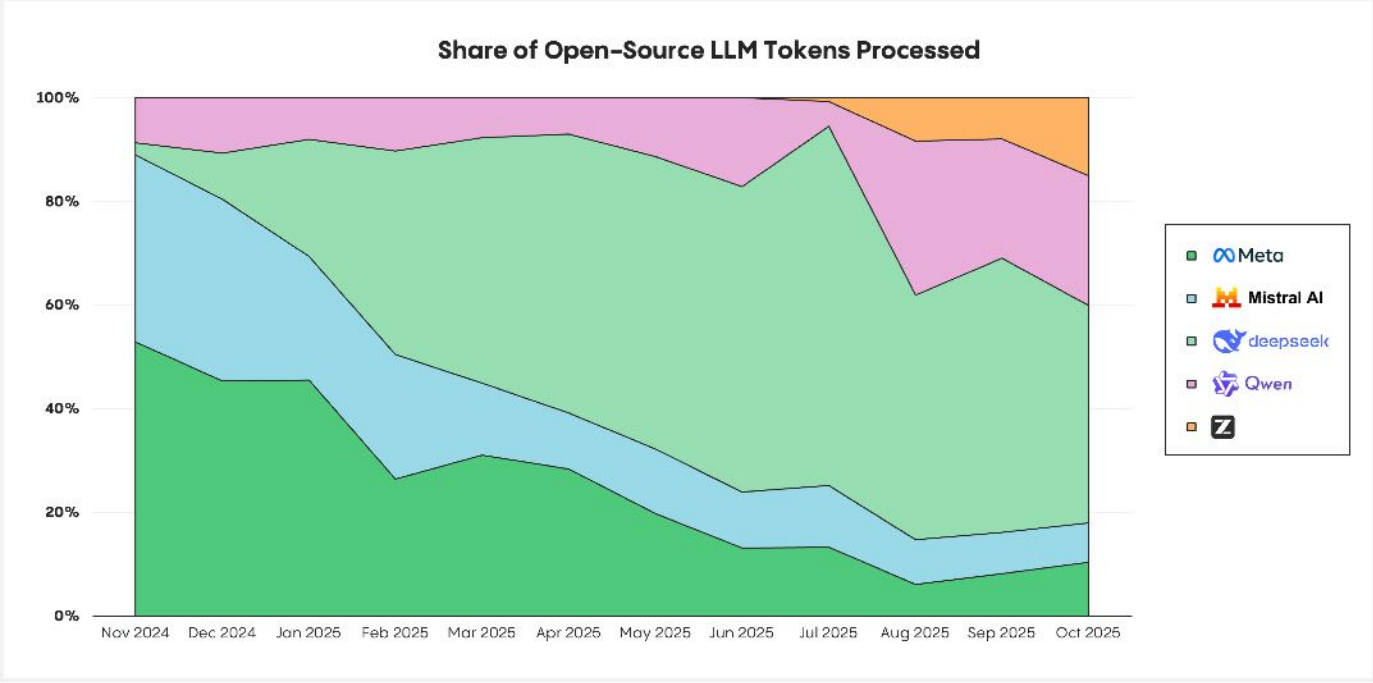
Despite falling off the frontier pace this year, Llama remains the most widely adopted open-weight model in the enterprise. But the model's stagnation—including no new major releases since the April release of Llama 4—has contributed to a decline in overall enterprise open-source share from **19%** last year to **11%** today.

Enterprises remain particularly cautious toward Chinese open-source models, despite their impressive progress this year and growing popularity among startups. Collectively, they account for just **1%** of total LLM API usage (roughly **10%** of enterprise open-source).

Outside the enterprise, adoption looks very different. vLLM and OpenRouter*,[13] two popular benchmarks for startup and indie developer usage, show rapidly rising adoption for Qwen, DeepSeek (V3, R1), Moonshot/Kimi, MiniMax, and Z AI's GLM, though DeepSeek's usage has moderated after an initial surge following its R1 launch.

Smaller models Qwen3 and GLM are especially popular for their competitive performance against much larger alternatives. Airbnb, for instance, relies on Qwen heavily for its user-facing AI features,[14] while Cursor uses the model as the open-source base for its internal model.[15]

Chinese Models Seeing Rising Adoption in the Broader Developer Ecosystem



The distribution of open-source LLM tokens continues to evolve, with Chinese models gaining visible traction among developers over the past year.

13. Based on publicly available adoption indicators from vLLM (https://github.com/vllm-project/vllm) and model usage share from OpenRouter (https://openrouter.ai), accessed November–December 2025.
14. Bloomberg, "Chesky Says OpenAI Tools Not Ready for ChatGPT Tie-Up With Airbnb App," October 21, 2025, https://www.bloomberg.com/news/articles/2025-10-21/airbnb-ceo-brian-chesky-says-chatgpt-integration-not-ready-for-airbnb-app
15. KrASIA, "Coding tools Cursor and Windsurf found using Chinese AI in latest releases," November 6, 2025, https://kr-asia.com/coding-tools-cursor-and-windsurf-found-using-chinese-ai-in-latest-releases; Al Jazeera, "China's AI is quietly making big inroads in Silicon Valley," November 13, 2025, https://www.aljazeera.com/economy/2025/11/13/chinas-ai-is-quietly-making-big-inroads-in-silicon-valley

AI Infrastructure: A Modern AI Stack Still in Development

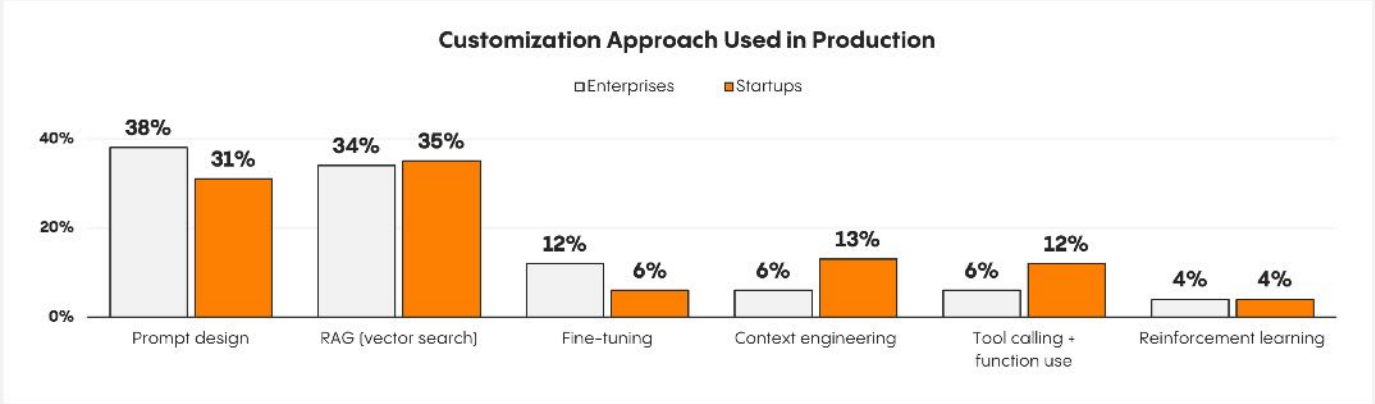
For all the talk of "agents," real production architectures remain surprisingly simple: Only **16%** of enterprise and **27%** of startup deployments qualify as true agents—systems where an LLM plans and executes actions, observes feedback, and adapts its behavior—while most are still built around fixed-sequence or routing-based workflows wrapped around a single model call.Customization patterns reinforce this technical nascency. **Prompt design** remains the dominant technique, followed by **retrieval-augmented generation** (RAG). More advanced approaches—**fine-tuning**, **tool calling**, **context engineering**, and **reinforcement learning** (RL)—are still niche and used primarily by frontier teams.

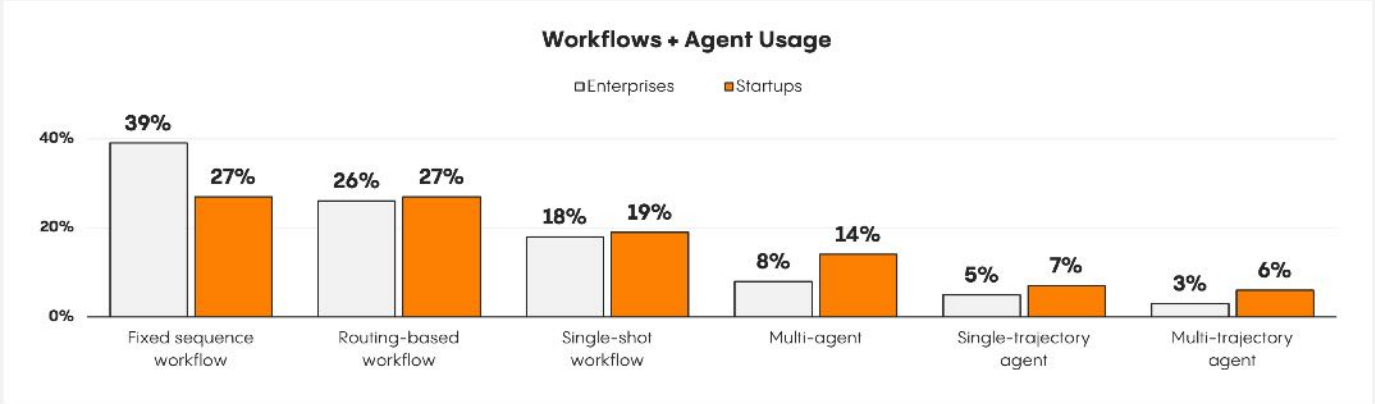
Because LLM-based application architectures continue to evolve gradually, the modern AI stack looks broadly similar to last year's. The biggest beneficiaries so far are incumbents extending trusted data and infrastructure platforms: Databricks, Snowflake, MongoDB, and Datadog.

Startup activity, on the other hand, clusters around inference and compute, where AI-native vendors compete directly with hyperscaler developer platforms. Inference platforms like Fireworks, Baseten, Modal, and Together win on performance and developer experience—offering serverless, high-throughput,

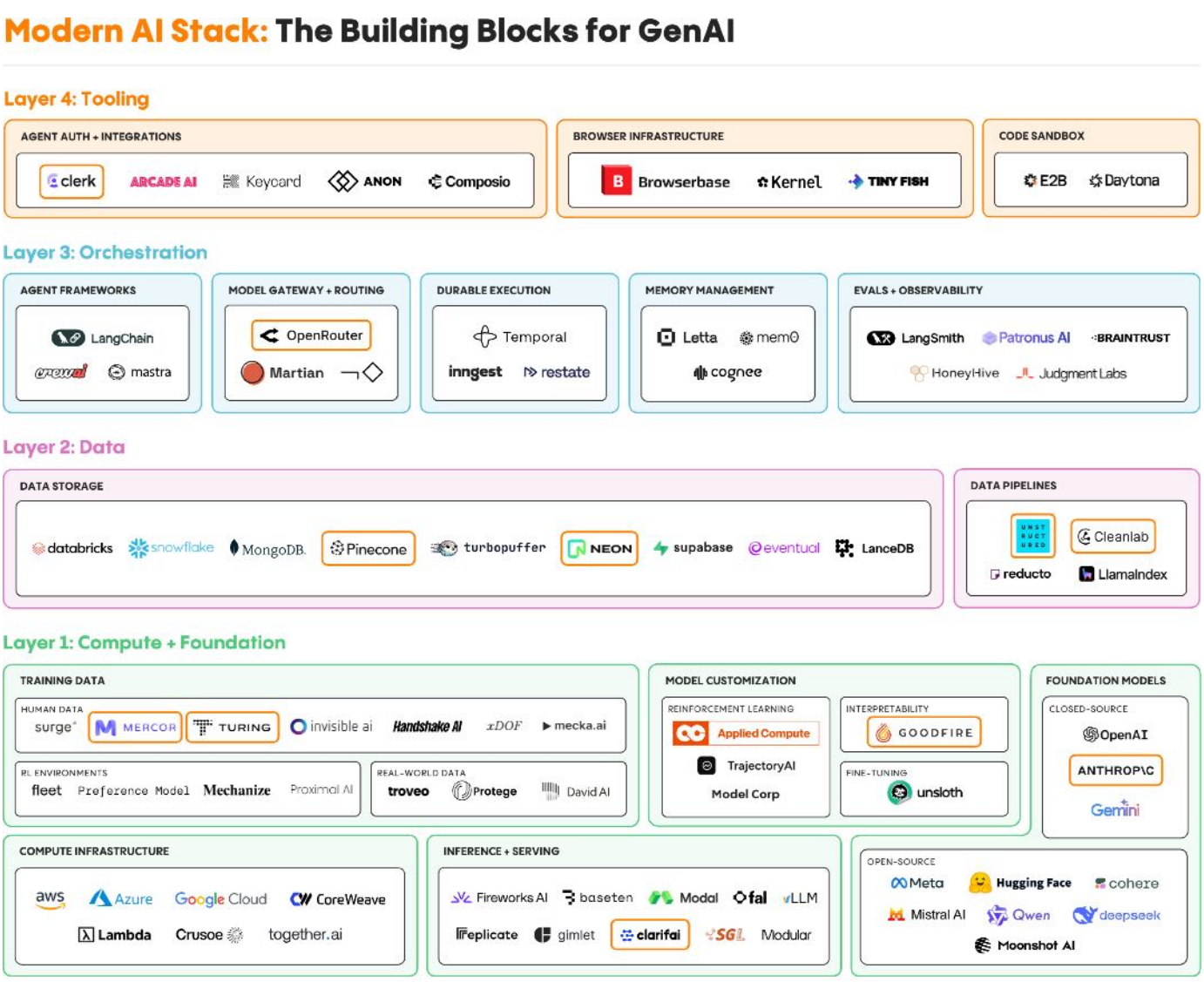
AI Architectures Remain Surprisingly Simple—Even in Production

Only 16% of enterprise deployments qualify as true agentic systems





Strip away the hype and most "AI agents" are basic if-then logic around a model call. Simple architecture works for today's use cases but reveals how early we are.



Enterprises invested $18B in AI infrastructure across foundation models, training systems, and the data and orchestration layers. This market map shows the companies serving each part of that stack.

open-weight endpoints with **2x+** speedups through handwritten or fused kernels, optimized serving stacks, and tightly managed GPU fleets.

Further up the stack, a new wave of observability and tooling vendors, including LangChain, Braintrust, and Judgment Labs, are building the runtime observability layer for AI, wedging in via developmental workflows including evals, tracing, and continuous learning. The market map below shows the key players building the foundational layers that power generative AI applications.

What's Next? Predictions for 2026

In 2025, AI became the fastest-scaling software category in history. Based on what we're seeing across the ecosystem, we have five predictions for the year ahead.

1. **AI will exceed human performance in daily practical programming tasks.**
   There is no plateauing of LLM skill sets, especially in verifiable domains such as math and programming, where the best models will continue to get better and better.

2. **Jevon's paradox continues to hold true.**
   Net spend on generative AI continues to rise despite falling costs of inference driven by orders-of-magnitude increase in inference volume.

   - Benchmarks continue to saturate, but will fail to completely capture real-world efficacy of models. Benchmark-maxxing models will not retain users long-term.
   - For frontier use cases like coding, users are actually quite price-insensitive and will pay more for performance.
   - Models gain widespread adoption for one big use case outside of programming.

3. **Explainability and governance go mainstream.**
   With the increase in autonomy and decision-making by agents, the ability to explain and govern the decisions they're making will increase in importance, driven by demand from the very people using the AI. We expect governments to ask for explainable decision-making and audit logs from agentic outcomes. Companies such as Goodfire*, which make neural networks interpretable and steerable, will become increasingly important to the enterprise.

4. **Models finally move to the edge.**
   Motivated by low-latency requirements, privacy/security, and other factors, compute will continue to move on-device, with the price of more and more non-frontier models approaching **$0**. Mobile manufacturers like Google, Apple, and Samsung will ship dedicated low-power GPU compute that delivers fast inference with no network and no cost on your phone.

Final Thoughts

Two years ago, when generative AI was still largely confined to pilots and proofs of concept, we published our first State of Generative AI in the Enterprise report. We set out to put real numbers behind what was happening, drawing from actual enterprise buyers versus analyst forecasts and vendor projections. We wanted to find the signal in the noise.

This year's findings make clear that the shift is no longer speculative. Enterprise AI is now a **$37 billion** market—the fastest-scaling category in software history. Across industries, AI has become core to how work gets done. Enterprises, seeing real returns, are doubling down.

We're fortunate to partner with many of the companies driving this shift: the frontier model provider leading the coding transformation, security platforms protecting enterprise AI at scale, and vertical applications redefining healthcare, legal, finance, and education. These teams are setting new standards and building the foundation for the next wave of innovation.

We are three years into this wholesale transformation. It's still early, but the first waves of leaders are emerging, and the value is clear. If you're a founder building at the frontier, we'd love to meet you.

Menlo Ventures is ALL IN on AI. Let's build what comes next together.

*Menlo Ventures investment

Data Sources and Methodology

Survey Approach

This report synthesizes findings from a survey of 495 U.S. enterprise AI decision-makers, conducted in partnership with an independent research firm, November 7–25, 2025. Respondents included C-suite executives, VPs of Engineering and Product, and technical leaders responsible for AI purchasing and development decisions at companies actively using AI tools.

Market Sizing Model

Our market sizing combines survey data from enterprise AI decision-makers with analysis of the generative AI ecosystem. We categorize companies by sector, type (startup vs. incumbent), and go-to-market motion (PLG vs. enterprise sales), drawing from publicly available information, industry reporting, and market analysis to estimate revenue distribution across the AI landscape.

Scope

Generative AI spending includes foundation models, model training infrastructure, AI infrastructure, and AI applications from both startups and incumbents. It excludes chips (e.g., Nvidia), inference and model serving (e.g., AWS, GCP, Azure, Fireworks), and AI features built into existing software solutions (e.g., Intuit Assist).

LLM Market Share

LLM market shares represent estimated dollars spent based on proportion of production API usage. Survey respondents reported the share of their AI workloads using each model. Responses were weighted based on each enterprise and startup application's scale, and triangulated with publicly reported financials where available.

Limitations

Market estimates represent our best assessment as of December 2025. The survey sample is limited to U.S. enterprises. Revenue estimates for private companies are based on public data and industry analysis.