Enterprise Agents Move From Demo to Deployment

AI News Digest

2026-04-23

Enterprise Agents Move From Demo to Deployment

By AI News Digest • April 23, 2026

The clearest shift today was enterprise AI moving toward governed, persistent agents: OpenAI launched workspace agents, Google expanded its agent platform and TPU stack, and Microsoft pushed agent mode deeper into Office while adding secure hosted sandboxes. Anthropic, Perplexity, and Sony AI added notable signals on labor, production post-training, and robotics.

Today's dominant theme: agents become enterprise infrastructure

OpenAI launches workspace agents for team workflows

OpenAI introduced workspace agents in ChatGPT: shared agents that can handle complex tasks and long-running workflows across tools and teams without constant supervision [1, 2]. They can pull context from docs, email, chats, code, and other systems; take approved actions such as updating Linear issues, creating docs, or sending messages; work inside Slack threads; and keep running in the background or on a schedule [3]. OpenAI says teams can build an agent once and share it across teams, and the feature is now in research preview for ChatGPT Business, Enterprise, Edu, and Teachers plans [4, 5].

Why it matters: OpenAI is positioning agents as shared operational tools for teams, not just one-user chat features.

Google couples an enterprise agent platform with new TPU infrastructure

At Cloud Next, Google said customer API traffic has risen to more than 16 billion tokens per minute, up from 10 billion last quarter, while launching the

Gemini Enterprise Agent Platform and a new "mission control" layer to build, scale, govern, and optimize agents [6]. Google DeepMind described the platform as an evolution of Vertex AI that brings together model selection, agent building, integration, security, and access to 200+ models through Model Garden, including Gemini 3.1 Pro, Gemini 3.1 Flash Image, Lyria 3, and Gemma 4 [7, 8]. Google also announced two eighth-generation TPUs—TPU8 for large-scale pre-training and TPU8i for post-training and inference—with Google saying they will also be available to Cloud customers by year-end [9].

Why it matters: Google is packaging a full enterprise stack—models, governance, and specialized infrastructure—around agent deployment.

Microsoft pushes agents into Office, secure sandboxes, and national-scale capacity

Microsoft said Agent Mode is now generally available and the default across Copilot in Word, Excel, and PowerPoint, calling it a big change to the Copilot experience [10]. Nadella said agents can reason over the "canvas" of work, including the spatial structure of spreadsheets, while Microsoft also introduced Hosted agents in Foundry, where each agent gets a dedicated enterprise-grade sandbox with durable state, built-in identity, and governance [10, 11]. Separately, Microsoft committed A$25 billion—its largest investment in Australia to date—to expand AI and cloud capacity, strengthen cybersecurity, and build digital skills [12].

> "Every agent will need its own computer." [11]

Why it matters: Microsoft is making agents more default inside productivity software while building the execution and regional infrastructure underneath them.

Other signals worth tracking

Anthropic turns its 81,000-user study into a recurring labor signal

Anthropic said its latest research on responses from nearly 81,000 Claude users found that workers in both the highest- and lowest-paid occupations reported the largest productivity gains from AI, but those with the biggest speedups also expressed the greatest concern about job displacement [13, 14]. It also said occupations with high Claude usage, such as software engineering, were more worried about displacement than lower-exposure roles [15]. To keep tracking these effects, Anthropic launched a monthly Economic Index Survey asking Claude users how AI is changing their work [16].

Why it matters: The same groups reporting the biggest upside are also among the most concerned about substitution, which makes this a useful ongoing indicator of how adoption is landing in practice.

Perplexity says post-training is now serving production traffic

Perplexity published details on an SFT + RL pipeline for accurate search-augmented answers, saying it improves search, citation quality, instruction following, and efficiency [17]. Aravind Srinivas said a Qwen-based model from this pipeline is Pareto-optimal on accuracy-cost curves, combines search and tool calls in one model, performs better than GPT and Sonnet on cost-efficiency for Perplexity's production queries, and is already serving a significant chunk of daily traffic [18].

Why it matters: This is a concrete case of post-training moving from benchmark talk into live product economics.

Sony AI's Ace robot reaches expert-level table tennis

Sony AI said its Ace project tackled a 40-plus-year unsolved problem by building a robot that can rally at full speed with elite human table-tennis players, with the work accepted for publication in *Nature* and featured on the cover [19]. hardmaru said the system uses reinforcement learning and Sony vision sensors to achieve expert-level play, calling it a big step for adaptive robotics [20].

Why it matters: It is a notable example of modern AI methods producing fast, adaptive control in a demanding physical task.

Sources

1. X post by @OpenAI
2. X post by @OpenAI
3. X post by @OpenAI
4. X post by @OpenAI
5. X post by @OpenAI
6. X post by @sundarpichai
7. X post by @GoogleDeepMind
8. X post by @GoogleDeepMind
9. How Google's 8th Generation TPUs Power the Agentic Era
10. X post by @satyanadella
11. X post by @satyanadella
12. X post by @satyanadella
13. X post by @AnthropicAI
14. X post by @AnthropicAI
15. X post by @AnthropicAI
16. X post by @AnthropicAI
17. X post by @perplexity_ai
18. X post by @AravSrinivas
19. X post by @SonyAI_global
20. X post by @hardmaru