# OpenAI unveils ‘ChatGPT agent’ that gives ChatGPT its own computer to autonomously use your email and web apps, download and create files for you

Want smarter insights in your inbox? Sign up for our weekly newsletters to get only what matters to enterprise AI, data, and security leaders. Subscribe Now  
  
OpenAI isn’t letting the delay of its open source AI model slow it down on shipping other features.  
  
Today, the company is unveiling ChatGPT agent, a feature that allows its AI chatbot to autonomously browse the web, conduct extensive research, download and create new files for its human users using its own virtual computer.  
  
Come again? ChatGPT now gets its own PC? And it can use that PC to log into your, the human user’s, accounts and download or send stuff for you?  
  
That’s correct, at least in a virtual sense, according to OpenAI. As the company explains:  
  
The AI Impact Series Returns to San Francisco - August 5  
  
The next phase of AI is here - are you ready? Join leaders from Block, GSK, and SAP for an exclusive look at how autonomous agents are reshaping enterprise workflows - from real-time decision-making to end-to-end automation.  
  
Secure your spot now - space is limited: https://bit.ly/3GuuPLF  
  
“The model can choose to open a page using the text browser or visual browser, download a file from the web, manipulate it by running a command in the terminal, and then view the output back in the visual browser. The model adapts its approach to carry out tasks with speed, accuracy, and eﬃciency.”  
  
Users can engage the agent by clicking on the ‘Tools’ button in the ChatGPT prompt entry box, opening the dropdown menu, and selecting ‘agent mode’ from the available options.  
  
Then, when it’s turned on describe a task in plain language, and the agent can carry it out across web and local app environments, combining reasoning with actions that previously only a human user could perform on their own machine, manually.  
  
ChatGPT agent can connect to apps like your personal or business Gmail and GitHub, so it can pull in useful information — emails or code — from your accounts to help with tasks you ask it to do. It can connect to third-party application programming interfaces (API) to pull information and use connected applications and services through them, as well.  
  
If a website needs you to log in, you can do that securely through a special browser view, which lets the agent dig deeper and handle more personalized tasks, like checking your inbox or filling out forms on your behalf.  
  
The new ChatGPT agent builds upon and expands from the “Operator” agent OpenAI released in January 2025, which allowed ChatGPT to browse the web and fill out forms, place orders, and do other web-based tasks in a private “headless browser,” that is, a cloud-based custom web browser that OpenAI itself maintained and offered for each Operator session.  
  
However, Operator was limited only to interacting with websites and web-based applications — not programs that could also run locally on a PC, such as spreadsheet tabulators and slide deck presentation making software.  
  
The new ChatGPT agent can browse websites, interact with online forms, run code, analyze data, and deliver finished outputs — such as editable presentations or spreadsheets — based entirely on user instructions.  
  
The unveiling comes on the heels of a report published days ago by independent subscription tech industry website The Information suggesting that OpenAI would upgrade ChatGPT to be a more direct competitor to its own investor Microsoft’s Office software applications (e.g. Excel, Word, PowerPoint, etc.)  
  
In fact, OpenAI positions ChatGPT agent as a merging of two of its prior agents — Operator and Deep Research, the latter introduced in February 2025, which exhaustively searches the web through its own headless text-only browser to find and compile information into lengthy and in-depth reports (hence the name). As OpenAI writes in a blog post:  
  
“Operator couldn’t dive deep into analysis or write detailed reports, and deep research couldn’t interact with websites to refine results or access content requiring user authentication. In fact, we saw that many queries users attempted with Operator were actually better suited for deep research, so we brought the best of both together.”  
  
The previous Operator tool will be phased out, but users can still access Deep Research via the dropdown in the ChatGPT interface.  
  
Whether using a visual browser to interact with a website or a terminal to run Python code, the agent moves seamlessly between tools within a single session.  
  
It supports a broad range of use cases, from analyzing competitors and generating reports to planning trips, summarizing emails, or booking appointments.  
  
Users can interrupt, redirect, or pause a task at any time, with the agent picking up right where it left off.  
  
Starting today, subscribers to ChatGPT’s $200-per-month “Pro” tier will have full access to ChatGPT agent, with a monthly quota of 400 messages.  
  
ChatGPT Plus ($20 per month) and Team ($30 per month) will gain access over the next few days, with 40 messages per month. Additional usage is available through credit-based options.  
  
OpenAI said in a release shared with VentureBeat under embargo that its ChatGPT Enterprise and Education subscribers will gain access to the feature the coming weeks.  
  
For now, the feature is not yet available in Europe or Switzerland, no doubt disappointing residents there.  
  
Given that the agent can now take actions on behalf of users — including on logged-in websites or with connected apps — OpenAI has introduced extensive safety measures.  
  
These include user confirmations before taking action, active supervision for sensitive tasks, and technical safeguards to limit unintended behavior.  
  
Key protections include:  
  
In line with its Preparedness Framework, OpenAI is treating ChatGPT agent as a High capability system in the biological and chemical domains.  
  
While there is no direct evidence of misuse, the company is activating its strongest safety safeguards out of caution.  
  
These include enhanced refusal training, red teaming by biosafety experts, and improved detection systems.  
  
Recall that Anthropic‘s recent information released about its new Claude Opus 4, and other surveys of advanced AI models, have shown that when given access to external tools and applications such as email, they can in some cases take actions they believe to be moral and ethical but that may compromise the user, such as emailing government agencies or journalists of suspected wrongdoing on the part of the user.  
  
The model thinks it is acting like a “whistleblower” but in fact, may compromise user privacy, security, and proprietary information and alert authorities to wrongdoing where there is none, or where it is dubious.  
  
ChatGPT agent isn’t just more capable in theory—it has delivered strong results across a number of benchmarks designed to simulate real-world knowledge work. It set a new high score on Humanity’s Last Exam with a 44.4 using parallel rollout methods, and achieved 27.4% on the difficult FrontierMath benchmark.  
  
On SpreadsheetBench, it scored 45.5%—more than doubling Copilot in Excel’s performance.  
  
Some features, like slideshow generation, are still in beta and may feel basic in formatting or differ slightly between in-app previews and exported files.  
  
OpenAI is actively training the next iteration of this feature to improve polish and layout.  
  
The launch of ChatGPT agent signals a shift in how users interact with AI—from asking questions to assigning complete tasks.  
  
With its ability to reason, act, and produce deliverables, OpenAI is betting that users increasingly want AI not just to assist them, but to work for them. While the company emphasizes that the agent is still evolving, it sees this launch as the foundation for a more interactive, action-oriented future for AI.  
  
If you want to impress your boss, VB Daily has you covered. We give you the inside scoop on what companies are doing with generative AI, from regulatory shifts to practical deployments, so you can share insights for maximum ROI.  
  
Read our Privacy Policy  
  
Thanks for subscribing. Check out more VB newsletters here.  
  
An error occured.

# Blaxel raises $7.3M seed round to build ‘AWS for AI agents’ after processing billions of agent requests

Want smarter insights in your inbox? Sign up for our weekly newsletters to get only what matters to enterprise AI, data, and security leaders. Subscribe Now  
  
Blaxel, a startup building cloud infrastructure specifically designed for artificial intelligence agents, has raised $7.3 million in seed funding led by First Round Capital, the company announced Thursday. The financing comes just one month after the six-founder team graduated from Y Combinator’s Spring 2025 batch, underscoring investor appetite for infrastructure plays in the rapidly expanding AI agent market.  
  
The San Francisco-based company is betting that the current generation of cloud providers — Amazon Web Services, Google Cloud, and Microsoft Azure — are fundamentally mismatched for the new wave of autonomous AI systems that can take actions without human intervention. These AI agents, which handle everything from managing calendars to generating code, require dramatically different infrastructure than traditional web applications built for human users.  
  
“The current cloud providers have been designed for the Web 2.0, Software as a Service era,” said Paul Sinaï, Blaxel’s co-founder and CEO, in an exclusive interview with VentureBeat. “But with this new wave of agentic AI, we believe that there is a need for a new type of infrastructure which is dedicated to AI agents.”  
  
The timing reflects a broader shift in enterprise computing as companies increasingly deploy AI agents for customer service, data processing, and workflow automation. Unlike traditional applications where databases sit alongside web servers in predictable patterns, AI agents create unique networking challenges by connecting to language models in one region, APIs in another cloud, and knowledge bases elsewhere—all while users expect instant responses.  
  
The AI Impact Series Returns to San Francisco - August 5  
  
The next phase of AI is here - are you ready? Join leaders from Block, GSK, and SAP for an exclusive look at how autonomous agents are reshaping enterprise workflows - from real-time decision-making to end-to-end automation.  
  
Secure your spot now - space is limited: https://bit.ly/3GuuPLF  
  
Blaxel has already demonstrated significant traction, processing millions of agent requests daily across 16 global regions by the end of their Y Combinator batch. One customer is running over 1 billion seconds of agent runtime to process millions of videos, representing a scale that illustrates the infrastructure demands of AI-first companies.  
  
“One of our customers is processing session replays to enable product managers to understand better how the user behavior of their product,” Sinaï explained. “They need to process millions of session replays every month. So it represents millions of minutes of sessions. They are using our agentic infrastructure to process those session replays and provide insights for product managers.”  
  
The company’s approach centers on providing infrastructure that AI agents can operate themselves, rather than requiring human administrators. This includes sandboxed virtual machines that boot in under 25 milliseconds, automatic scaling based on agent activity patterns, and APIs designed to be consumed directly by AI systems rather than human developers.  
  
Blaxel’s unusual six-founder structure stems from the team’s shared experience building and selling a previous company to OVHcloud, Europe’s largest cloud provider. That company became OVH’s entire analytics product suite, giving the team firsthand experience with both cloud infrastructure challenges and successful exits.  
  
“I know it sounds unusual, pretty big team. We didn’t fit exactly on the stage for demo day,” Sinaï said, referencing Y Combinator’s signature event. “But we already did that. My previous company, which I sold to OVH cloud, we were also six co-founders.”  
  
The team includes Charles Drappier, whom Sinaï has known for over 15 years, along with co-founders Christophe Ploujoux, Nicolas Lecomte, Thomas Crochet, and Mathis Joffre. Their collective experience spans infrastructure, developer tools, and platform engineering — critical expertise for competing against tech giants with virtually unlimited resources.  
  
“I think it’s important to be six right now, because we have a lot of ambition,” Sinaï said. “What we are doing is building this next generation of cloud computing for this new agentic era.”  
  
The cloud infrastructure market is notoriously competitive, with AWS commanding roughly one-third market share and newer players like Modal, Replicate, and RunPod targeting AI workloads. Blaxel differentiates itself by focusing specifically on AI agents rather than model inference or training.  
  
“Most of the competitors you mentioned are solving a very difficult problem, which is around the inference — how you can host your model, how you can make those models as fast as you can in terms of number of tokens,” Sinaï said. “But there is not that many people working on infrastructure for the agents, and it’s exactly what we are doing.”  
  
The company’s platform includes three main components: agent hosting for deploying AI systems as serverless APIs, MCP (Model Context Protocol) servers for connecting agents to external tools, and a unified gateway for accessing multiple AI models. The infrastructure is designed to handle the variable resource demands of AI agents, which might require minimal computing power while waiting for responses but need significant resources during active processing.  
  
Despite targeting younger AI-first companies, Blaxel has implemented enterprise-grade security measures including SOC2 and HIPAA compliance. The platform offers data residency controls that allow customers to restrict workloads to specific geographic regions—critical for companies in regulated industries.  
  
“We provide a policy framework where you can attach, for example, to workloads to say, this agent cannot run outside of those subsets of regions,” Sinaï explained. “You can attach a policy to say this agent cannot run outside of the United States, so you are sure that this agent will process the data only in the regions you have chosen.”  
  
This approach reflects the company’s belief that even early-stage AI companies need robust infrastructure practices because they’re building the enterprises of tomorrow. “We believe that it’s very important to have, even for young companies, the best infrastructure with the best practices, because they are going to become enterprises,” Sinaï said.  
  
Blaxel has adopted a pay-as-you-go pricing model similar to established cloud providers, moving away from an initial subscription approach after validating market demand during their Y Combinator batch. The model charges customers only when their agents are actively processing tasks, shutting down infrastructure during idle periods to optimize costs.  
  
“We provide infrastructure that spin up in just few milliseconds and shut down in just one second,” Sinaï said. “So you just pay for the time your agent is actually processing something. When your agent is waiting for something else, you don’t have to pay for it because we shut it down.”  
  
The approach has already delivered cost savings for customers, with one client achieving 50% cost reduction compared to typical serverless solutions while processing terabytes of data monthly.  
  
The investment comes as industry analysts predict explosive growth in AI agent adoption. Gartner forecasts that 75% of application development will involve AI agents by 2028, though Sinaï believes current enterprise adoption remains largely experimental.  
  
“Right now, most of companies working actively in production are mostly smaller companies, not yet enterprise companies,” he said. “So we are focusing really on serving them exactly like the big cloud providers did in the past.”  
  
The strategy mirrors how Amazon Web Services initially focused on startups and developer-friendly companies before expanding to enterprise customers. Blaxel plans to follow a similar path, using the $7.3 million to expand their software platform before potentially moving into custom hardware and data center optimization.  
  
“Seven millions is not enough to build data centers, obviously, but I think it’s important to go step by step,” Sinaï said. “Being sure that right now we have the best interfaces we can provide to our customers, the best services for their agents, and then going into the deeper infrastructure optimization.”  
  
The company’s roadmap includes features like snapshot forking for agent experimentation, automatic failover capabilities, and deeper optimization for the massive scale they anticipate. With projections of hundreds of billions of AI agents in the coming decades, Blaxel sees an opportunity to build infrastructure designed for this new computing paradigm from the ground up.  
  
“We believe that there is a huge economy which is starting around the agents,” Sinaï said. “There are going to be hundreds of billions of AI agents, and the infrastructure we have today has not been designed for this new wave.”  
  
The funding round included participation from Y Combinator, Liquid2, Transpose, and angel investors who share the company’s vision of purpose-built agent infrastructure. As AI agents transition from experimental tools to production systems handling critical business processes, Blaxel’s specialized approach could position it to capture significant market share in what may become the next major category of cloud computing.  
  
If you want to impress your boss, VB Daily has you covered. We give you the inside scoop on what companies are doing with generative AI, from regulatory shifts to practical deployments, so you can share insights for maximum ROI.  
  
Read our Privacy Policy  
  
Thanks for subscribing. Check out more VB newsletters here.  
  
An error occured.

# AWS unveils Bedrock AgentCore, a new platform for building enterprise AI agents with open source frameworks and tools

Want smarter insights in your inbox? Sign up for our weekly newsletters to get only what matters to enterprise AI, data, and security leaders. Subscribe Now  
  
Cloud giant Amazon Web Services (AWS) believes AI agents will change how we all work and interact with information, and that enterprises need a platform that allows them to build and deploy agents at scale — all in one place.  
  
Today at its New York Summit, AWS unveiled Amazon Bedrock AgentCore, a new enterprise-grade platform designed to build, deploy, and operate AI agents securely and at scale.  
  
Swami Sivasubramanian, AWS Vice President of Agentic AI, said during the keynote that AgentCore “helps organizations move beyond experiments to production-ready agent systems that can be trusted with your most critical business processes.”  
  
AgentCore is a modular stack of services—available in preview—that gives developers the core infrastructure needed to move AI agents from prototype to production, including runtime, memory, identity, observability, API integration, and tools for web browsing and code execution.  
  
The AI Impact Series Returns to San Francisco - August 5  
  
The next phase of AI is here - are you ready? Join leaders from Block, GSK, and SAP for an exclusive look at how autonomous agents are reshaping enterprise workflows - from real-time decision-making to end-to-end automation.  
  
Secure your spot now - space is limited: https://bit.ly/3GuuPLF  
  
“We believe that agents are going to fundamentally change how we use tools and the internet,” said Deepak Singh, AWS Vice President of Databases and AI. “The line between an agent and an application is getting blurrier.”  
  
AgentCore builds on the existing Bedrock Agents framework, launched in late 2024, but dramatically expands capabilities by supporting any agent framework or foundation model—not just those hosted within Bedrock.  
  
That includes compatibility with open-source toolkits like CrewAI, LangChain, LlamaIndex, LangGraph, and AWS’s own Strands Agents SDK.  
  
AgentCore also integrates with the AWS Marketplace, enabling teams to discover and deploy pre-built agents and tools.  
  
According to Singh, AgentCore has been designed with interoperability in mind. It supports emerging industry standards like MCP and Google’s Agent-2-Agent (A2A) protocol. Features such as AgentCore Identity and Gateway ensure agents have clear permissioning and can interact securely with internal systems and third-party APIs.  
  
AWS’s launch puts it squarely into the center of what’s quickly becoming one of the most competitive segments in enterprise AI.  
  
OpenAI’s Agents SDK and Google’s Gemini-based Agents SDK are both pushing similar visions of end-to-end agent development platforms.  
  
Writer’s AI HQ and startups like Cognition (maker of Devin) are also building tools for managing autonomous software agents.  
  
“Agents are the most impactful change we’ve seen in ages,” Sivasubramanian said. “With agents comes a shift to service as a software. This is a tectonic change in how software is built, deployed and operated.”  
  
Several companies granted early access to AgentCore are already building production-grade agentic applications across industries including finance, healthcare, marketing, and content management.  
  
Cloud document and file storage company Box is exploring ways to extend its content management tools using Strands Agents and Bedrock AgentCore Runtime.  
  
CTO Ben Kus said the integration gives Box customers “top tier security and compliance” while scaling AI capabilities across enterprise environments.  
  
Brazil’s Itaú Unibanco is using AgentCore to support its development of hyper-personalized, secure digital banking experiences. Chief Technology Officer Carlos Eduardo Mazzei said the new platform “will help us deliver an intuitive banking experience with the efficiency of automation and personalization customers expect.”  
  
In the healthcare space, Innovaccer has built a new protocol—HMCP (Healthcare Model Context Protocol)—on top of AgentCore Gateway. CEO and co-founder Abhinav Shashank called Gateway a “game-changer” that allows the company to convert existing APIs into agent-compatible tools at scale while maintaining trust, compliance, and operational efficiency.  
  
Marketing firm Epsilon is leveraging AgentCore to accelerate campaign build times and improve engagement. Prashanth Athota, SVP of Software Engineering, said the company expects to reduce build times by up to 30% and enhance customer journey personalization.  
  
AgentCore is now available in preview in select AWS regions including US East (N. Virginia), US West (Oregon), Asia Pacific (Sydney), and Europe (Frankfurt). It’s free to try until September 16, 2025, with pricing to begin thereafter.  
  
Pricing for AgentCore is entirely consumption-based, with no upfront commitments or minimum fees. Each module—Runtime, Memory, Identity, Observability, Gateway, Browser, and Code Interpreter—is billed independently and can be used a la carte or together.  
  
Runtime, Browser, and Code Interpreter services are priced per second, based on CPU and memory usage, with rates set at $0.0895 per vCPU-hour and $0.00945 per GB-hour.  
  
Gateway charges $0.005 per 1,000 tool API invocations, $0.025 per 1,000 search queries, and $0.02 per 100 tools indexed per month.  
  
Memory costs are based on data volume: $0.25 per 1,000 short-term memory events, $0.75 per 1,000 long-term memories stored (or $0.25 with custom strategies), and $0.50 per 1,000 retrievals.  
  
AgentCore Identity costs $0.010 per 1,000 token or API key requests, though it’s included at no extra charge when used via Runtime or Gateway.  
  
Observability is billed via Amazon CloudWatch rates.  
  
To learn more or get started, AWS directs developers to its AgentCore documentation, GitHub samples, and a dedicated Discord server.  
  
If you want to impress your boss, VB Daily has you covered. We give you the inside scoop on what companies are doing with generative AI, from regulatory shifts to practical deployments, so you can share insights for maximum ROI.  
  
Read our Privacy Policy  
  
Thanks for subscribing. Check out more VB newsletters here.  
  
An error occured.